A chimera model for motion anticipation in the retina and the primary visual cortex

Jérôme Emonet, Selma Souihel, Matteo Di Volo, Alain Destexhe, Frédéric Chavane, Bruno Cessac

A CHIMERA MODEL FOR MOTION ANTICIPATION IN THE RETINA AND THE PRIMARY VISUAL CORTEX

Jerome Emonet
Université Côte d'Azur, Inria
Biovision team and Neuromod Institute
Sophia Antipolis, France
jerome.emonet@inria.fr

Selma Souihel
P16 - Programme IA
Inria Rocquencourt
Domaine de Voluceau
78150 Le Chesnay-Rocquencourt
selma.souihel@inria.fr

Matteo di Volo
Université Claude Bernard Lyon 1
Institut National de la Santé et de la Recherche Médicale,
Stem Cell and Brain Research Institute U1208
Bron, France
matteo.di-volo@univ-lyon1.fr

Alain Destexhe
Institut NeuroPSI - UMR9197
CNRS Paris-Saclay University
Campus CEA Saclay, Saclay, FRANCE
Alain.Destexhe@cnrs.fr

Frederic Chavane
Institut de Neurosciences de la Timone UMR7289
CNRS Aix-Marseille University
13385 Marseille Cedex 05, France
frederic.chavane@univ-amu.fr

Bruno Cessac
Université Côte d'Azur, Inria
Biovision team and Neuromod Institute
Sophia Antipolis, France
bruno.cessac@inria.fr

December 12, 2024

ABSTRACT

Motion anticipation refers to the capacity of the visual system to compensate for inherent delays in visual processing. This ability results from distinct mechanisms taking place in the retina (Berry et al., 1999) and in the visual cortex (Jancke et al., 2004). To study their respective role, we propose a mean field model of the primary visual cortex (V1) connected to a realistic retina model. Our first goal is to reproduce experimental results on motion anticipation, made in monkeys by using voltage dye optical imaging (VSDI) (Benvenuti et al., 2020), and to assess the impact of the retina in this process. For this, we first study the model in the case where the retina does not itself provide anticipation. Then, anticipation is only triggered by a cortical mechanism, called "anticipation by latency". As we show, this mechanism strongly depends on the intensity of the retinal input supplied to the cortex, even if the retina does not itself provide anticipation. We also unravel the effect of the stimulus features, such as speed and contrast, and report the impact of physiological parameters not accessible experimentally, such as the size of cortical extensions or fibre conduction. Then, we explore the changes in the cortical wave of anticipation when V1 is triggered by a retina output implementing different potential retina-driven anticipatory mechanisms, including gain control and lateral inhibition by amacrine cells. In this setting, we show how retinal and cortical anticipation combine, to provide an efficient processing where the VSDI signal response is in advance over the moving object that triggers this response, compensating the delays in visual processing, in full agreement with the experimental results of (Benvenuti et al., 2020).

***K*eywords :** Visual processing · Anticipation · Retina · V1 · Lateral connectivity

1 Introduction

The brain is not just a reactive encoder, it is able to respond in a proactive way, especially with regards to stimuli. In visual processing, it is able to extrapolate and estimate how the visual stimulus is the most likely to behave in a near future, given the information of the past. This is, of course, based on the assumption that the stimulus is predictable to some extent. In particular, several studies have suggested that the predictability of stimuli can be learned from spatial and temporal regularities, arising e.g. in a deterministic trajectory. However, one can ask at which stage of the visual system do these predictions start taking place. At the level of the early stages of the visual system ? Or in late-stage processors ?

Prediction is a rather wide concept. If by this term, one refers to motion extrapolation, studies have reported that it already starts at the level of the retina. Berry et al. (Berry et al., 1999) have first shown that local gain control mechanisms, occurring at the level of retinal bipolar and ganglion cells, can explain a form of local anticipation for a moving bar by advancing the peak in the retinal ganglion cells response. This explains the change in the shape of responses observed in experimental data - bringing the ganglion cells to their activity peak earlier than when they respond to a flashed bar - without modifying the time at which their activity starts increasing, i.e. when the bar enters in their receptive field. Another study by Johnston et al. (Johnston and Lagnado, 2015) has emphasized the role of lateral inhibition in the elicitation of anticipatory mechanisms at the retinal level.

They have proposed that motion anticipation can be mediated via feedforward inhibition from amacrine cells inputs that specifically suppress the response to the moving object in the latter half of the receptive field. This mechanism as well truncates the response and yields an early response peak. According to the authors, this "adaptation anticipation" occurs at the level of synapses. It requires each excitatory synapse to be more distal than an inhibitory synapse on the pathway. The latter would inhibit its EPSP and generate a shift feedforward (anticipation) of the peak. Thus, there must be an excess of inhibitory synapses compared to excitatory ones. In this case, motion anticipation arises from the general properties of the retina connectome and from the feedforward inhibition that ganglion cells receive from amacrine cells. Souihel and Cessac (Souihel and Cessac, 2021) explored, in a modelling study, another potential anticipatory effect of amacrine cells. In addition to the feed-worward effects discussed in (Johnston and Lagnado, 2015) feedback inhibition due to amacrine cells could induce a wave of activity further anticipating the bipolar cell response and thereby enhancing the effect of gain control. Finally, Menz et al (Menz et al., 2020) observed that motion anticipation can, in addition to a peak advancement by truncation of the response, arise by advancing the *onset* of the response, with a strong anticipatory effect. They hypothetize that this effect may come from amacrine cells: hyper-polarizing these cells would provide a dis-inhibitory input to ganglion cells prior to the object crossing the receptive field center.

It has also been shown that anticipation is further carried out at the level of the primary visual cortex (Jancke et al., 2004; Guo et al., 2007; Subramaniyan et al., 2018; Benvenuti et al., 2020). Jancke et al. (Jancke et al., 2004) first demonstrated the existence of anticipatory mechanisms in the cat primary visual cortex. They recorded cells in the central visual field of area 17 of anesthetized cats, responding to small squares of light, either flashed or moving in different directions, and with different speeds. When presented with the moving stimulus, these cells show a reduction of neural latencies, as compared to the flashed stimulus. Guo et al. (Guo et al., 2007) noted the ability of spatial regions, beyond the receptive field of V1 neurons, to influence the response of these neurons. A spatio-temporal sequence of bars, of the same orientation, flashed at the periphery of the receptive field of a V1 neuron, generates an increase in the neuron's activity if the orientation matches its preferred orientation. This phenomenon does not occur if the spatiotemporal sequence is randomised. Subramaniyan et al. (Subramaniyan et al., 2018) have reported the existence of similar anticipatory effects in the macaque primary visual cortex, showing that a moving bar is processed faster than a flashed bar. They gave two possible explanations to this phenomenon : either a shift in the cells receptive fields, induced by motion, or a faster propagation of motion signals. Consistent with the study by Jancke et al., they reported a speed dependence of the response latency as well as a luminance dependence. However, Subramaniyan et al. note that the motion representation delays are not reduced to zero, irrespective of the experimental setting of the flash lag effect. As a consequence, moving objects representation in V1 should be mislocalized. This observation is in favor of a collaborative work conducted, on the one hand, by the retina and V1 to help reducing the latencies, and, on the other hand, by other specialized brain regions which carry out predictive computations. Learning and training seem also to play important roles in calibrating the response of the nervous system to a giving moving object. Finally, Benvenuti et al. (Benvenuti et al., 2015, 2020) have

studied the trajectory-based activity in V1. They have compared voltage sensitive dye imaging responses for different cortical locations along a bar trajectory. By centering all positions on the same relative time events (bar centered in the middle of the receptive field), they have highlighted a gradient in the response (see also Fig. 2). The further the bar starts from the current position of a cortical column the earlier its activity rises. They gave convincing arguments that this increase is carried by the horizontal (intra-cortical) connectivity in the cortex.

At the current state of the art there are convincing benches of evidence that anticipation takes place along the pathway from retina to cortex. This has been observed with a large diversity of recording conditions (in vitro, in vivo, anesthetized, awake), for different animal models (rodent, cat, monkey . . .). This suggests that there might exist a synergy between the retina, the thalamus-LGN and the cortex to optimize anticipation. This leads to a natural question: How does the dynamical response of the retina to a moving object affect the dynamics changes in the cortex ?

Along these lines, the purpose of this paper is to understand what the retina brings to the primary visual cortex in terms of anticipation, using a computational model, grounded on experimental studies performed at the Institut des Neurosciences de la Timone, in F. Chavane Lab, as well as preexisting models themselves grounded on experiments. Rather than developing a biologically plausible model for one species, the aim of this article is to reconcile two sets of work that have been conducted to model different species, and to study the potential effect of known retinal anticipatory mechanisms on cortical anticipation. The model consists first of a retinal model developed in (Souihel and Cessac, 2021) to confront several potential mechanisms of retinal anticipation (gain control, amacrine cells) and further studied in (Cessac, 2022; Kartsaki et al., 2024). This retinal model is fed into a cortical mean-field model to compare what happens in the absence (non-biological case) or presence (biological case) of retinal anticipation. The cortical mean-field model of V1 has been previously developed by A. Destexhe and his collaborators (Zerlaut et al., 2018; Chemla et al., 2019), and calibrated on experimental VSDI signal data on the monkey cortex V1, performed in F. Chavane Lab. Note that, as further discussed in the text, we will not consider the effects of the thalamus (LGN) in this work. More precisely, it will be transparent, considered as a simple relay.

The paper is organized as follows. Section 2 is devoted to methods. We introduce the retino-cortical model in sections 2.1, 2.2 with some complementary elements added in the appendix, section 5.2. In section 2.3 we comment the originality of the model compared to existing work and explain the terminology "Chimera". In section 2.4 we introduce specific quantities (observables) used throughout the paper to quantify anticipation. Section 3 presents our main results. In section 3.1 we study the VSDI signal response to a moving bar when the retina is passive and brings no anticipatory effect (control conditions). We show that the model qualitatively reproduces the results experimentally observed in (Benvenuti et al., 2020) with quantitative deviations, presumably due to the lack of anticipation mechanisms coming from the retina. We analyse the effect of varying the retinal output amplitude (section 3.2.1) and show that it impacts the anticipation, even if the retina is passive. We also investigate the role of the stimulus speed (section 3.2.2), the cortical extent (section 3.2.3), and the conduction speed (section 3.2.4). In section

3.3 we analyse the effect of retinal anticipation mechanisms on cortical anticipation. The section 3.3.1 considers the impact of gain control applied either on bipolar cells or retinal ganglion cells on anticipation. In section 3.3.2 we analyse the effect of the amacrine cells connectivity. Here, we are interested on two types of connectivities: 1) feedforward connection from bipolar cells to amacrine cells to retinal ganglion cells, studied e.g. in (Johnston and Lagnado, 2015), which specifically suppresses the response to the moving object in the latter half of the receptive field truncating the response and yielding an early response peak; 2), feedback reciprocal connections between bipolar and amacrine cells, both imputing retinal ganglion cells. This allows us the compare the respective effects of these two mechanisms on the visual cortex anticipation. In section 4 we discuss several potential extensions of this work such as more complete models of the retina (including gap junctions or multiple amacrine cells connections), or the inclusion of a mean-field model of the thalamus. The appendix contains additional material concerning the model definition, its parameters and some results.

2 Methods

2.1 The retina model

The model is made of 3 bidimensional layers whose structure is shown in Fig. 1: A layer of bipolar cells (BCs), a layer of amacrine cells (ACs) and a layer of retinal ganglion cells (RGCs). The layers are rectangle grids, with the same dimension $L_x \times L_y\ mm^2$ and the same number of cells, N. BCs are labelled with an index $i = 1 \dots N$, ACs are labelled with an index $j = 1 \dots N$, RGCs with an index $k = 1 \dots N$. Cells are located on the nodes of their grid layer and are spaced by a distance δ mm in both directions x, y. Spatial coordinates are noted $\vec{x} = (x, y)$ and the coordinates of, e.g., BC i are (x_i, y_i).

The 3 layers ought to be parametrised by a vertical coordinate, z. However, this parametrisation is implicit in the retinal layers label BCs, ACs, RGCs and we are not going to use a vertical distance between layers. Since all layers have the same grid spacing, δ, and the same number of neurons, there is a vertical alignment of nodes: the BC with e.g. index $i = 10$ is vertically aligned with the AC of index $j = 10$ and the RGC with index $k = 10$. This given, we define the Euclidean distance between a cells i, in layer 1, and a cell j, in layer 2 as $d(i,j) = \sqrt{(\,x_i - x_j\,)^2 + \left(\,y_i - y_j^2\,\right)}$. That is, we do not consider the vertical distance, for simplicity. The distance $d(i,j)$ is used for connectivity patterns.

There is indeed an interlayer connectivity. The connectivity from BCs to ACs is characterized by a connectivity matrix Γ_A^B, with entries $\Gamma_{A_j}^{B_i} = 1$ if there is a connection from BC i to AC j, and $\Gamma_{A_j}^{B_i} = 0$ otherwise. In the paper, the connectivity structure of Γ_A^B is called "nearest-neighbours+1": the BC with coordinate i connects to the AC with coordinate i and with the four nearest neighbours of this AC. The synaptic weight from BC i to AC j is then $W_{A_j}^{B_i} = w_A^B\,\Gamma_{A_j}^{B_i}$ where the parameter $w_A^B \geq 0$ controls the excitatory synapses amplitude. This form of synaptic weigths allows us to tune the synaptic intensity from BCs to ACs with the unique parameter w_A^B. The connectivity from ACs to BCs is also characterized by a synaptic weight matrix W_B^A with entries

$W_{B_i}^{A_j} = w_B^A \Gamma_{B_i}^{A_j}$, $w_B^A \leq 0$ (inhibition from AC j to BC i) where Γ_B^A is "one to one" (AC j only connects to BC with index $i = j$). The synapse from BC i to RGC k corresponds to "Gaussian pooling" (Berry et al., 1999) where $W_{G_k}^{B_i} = w_G^B \frac{e^{-\frac{d(i,k)^2}{2\sigma^2}}}{2\pi\sigma^2}$ with $w_G^B \geq 0$. Likewise, the connectivity from AC j to RGC k is characterized by a synaptic weight $W_{G_k}^{A_j} = w_G^A \frac{e^{-\frac{d(j,k)^2}{2\sigma^2}}}{2\pi\sigma^2}$ with $w_G^A \leq 0$. Synaptic weights are expressed in Hz. The different types of connectivity are summarized in the appendix 5.3.

Cell types have characteristic times, expressed in seconds, corresponding to the integration time of their response to external influence and including synaptic and transmission delay. Here, we consider that all BCs have the same characteristic time, τ_B. Likewise, all ACs have a characteristic time, τ_A, and RGCs a characteristic time, τ_G.

The dynamics of cells is based on their voltage. We note V_{B_i} the voltage of BC i and so on. Voltage rectification takes place below a certain threshold (eq. (3)). In addition, BCs and RGCs have gain control, a desensitization when activated by a steady illumination. (eq. (5), (8)), characterized by an activation variable A_{B_i} for BCs, A_{G_k} for RGCs (Chen et al., 2013). The dynamics of voltages and activations is given by eq. (2) below.

BCs receive a visual input featuring the pre-processing of a visual stimulus via photo-receptors and horizontal cells. In this study, a visual stimulus is a grey scale video, that is a function $\mathcal{S}(\vec{x}, t) \in [\,0, 1\,]$ where 0 corresponds to black and 1 to white. The pre-processing of this visual stimulus via photo-receptors and horizontal cells is modelled by a spatio-temporal convolution:

$$\left[\mathcal{K}_{B_i} \overset{\vec{x},t}{*} \mathcal{S} \right](t) \equiv V_{i_{drive}}(t), \tag{1}$$

where $\mathcal{K}_{B_i}(\vec{x}, t)$ is a spatio-temporal kernel, centered at the coordinate of the BC i, and called "OPL kernel" (where "OPL" stands for "Outer Plexiform Layer" `https://www.ncbi.nlm.nih.gov/books/NBK11518/`). Spatially, this is a Gaussian with a center of radius σ_c and temporally, a gamma function with a characteristic time τ_c (see appendix 5.1 for the value of these parameters). Therefore, the OPL input of BCs is monophasic in space and time. As shown in (Kartsaki et al., 2024) the presence of lateral inhibition (here, by ACs) allows nevertheless to generate biphasic profiles in space and in time for the BCs response. Note that the spatial RF is circular: we do not consider orientation selective cells in this paper. In the definition of $\mathcal{K}_{B_i}$ there is a multiplicative factor, $\mathcal{C}$, which allows us to control the amplitude of $V_{i_{drive}}(t)$. This is used, in section 3.2.1, to modify the amplitude of the retinal input to the cortex.

The joint evolution of BCs, ACs, RGCs, driven by the stimulus $\mathcal{S}$ is given by the following set of equations. We use the standard notations of dynamical systems theory, where the time variable is omitted except for the non autonomous term (here, the drive term). We refer to the papers (Souihel and Cessac, 2021; Cessac, 2022; Kartsaki et al., 2024) for detail about this dynamical system and its mathematical study.

$$\left\{\begin{array}{lcl} \frac{dV_{B_i}}{dt} & = & -\frac{V_{B_i}}{\tau_B} + \sum_{j=1}^{N_A} W_{B_i}^{A_j} V_{A_j} + V_{i_{drive}}(t), \\ \frac{dA_{B_i}}{dt} & = & -\frac{A_{B_i}}{\tau_{a_B}} + h_B \mathcal{N}_B(V_{B_i}), \\ \frac{dV_{A_j}}{dt} & = & -\frac{V_{A_j}}{\tau_A} + \sum_{i=1}^{N_B} W_{A_j}^{B_i} R_B\left(V_{B_i}, A_{B_i}\right), \\ \frac{dV_{G_k}}{dt} & = & -\frac{V_{G_k}}{\tau_G} + \sum_{j=1}^{N_A} W_{G_k}^{A_j} V_{A_j} + \sum_{i=1}^{N_B} W_{G_k}^{B_i} R_B\left(V_{B_i}, A_{B_i}\right), \\ \frac{dA_{G_k}}{dt} & = & -\frac{A_{G_k}}{\tau_{a_G}} + h_G \mathcal{N}_G(V_{G_k}), \end{array}\right. \tag{2}$$

where, in addition to voltages, we have introduced the activity variables, A_{B_i} for BCs, A_{G_k} for RGCs, ruling the gain control mechanisms on these cell type. There is no gain control on ACs. BCs are, in addition, rectified. The function:

$$\mathcal{N}_B(V_{B_i}) = \left\{\begin{array}{ll} 0, & \text{if} \quad V_{B_i} \leq \theta_B; \\ V_{B_i} - \theta_B, & \text{else,} \end{array}\right. \tag{3}$$

models this BCs voltage rectification, where θ_B is the rectification threshold. The BCs output to ACs and RGCs is then characterized by a non linear response to its voltage variation, given by :

$$R_B\left(V_{B_i}, A_{B_i}\right) = \mathcal{N}_B\left(V_{B_i}\right)\, \mathcal{G}_B\left(A_{B_i}\right), \tag{4}$$

where the function:

$$\mathcal{G}_B(A) = \left\{\begin{array}{ll} 0, & \text{if } A \leq 0; \\ \frac{1}{1+A^6}, & \text{otherwise.} \end{array}\right. \tag{5}$$

implements the gain control of BCs as a function the activity variable A_{B_i} (Berry et al., 1999).

As ganglion cells are spiking cells, their response function is:

$$R_G\left(V_{G_k}, A_{G_k}\right) = \mathcal{N}_G(V_{G_k})\, \mathcal{G}_G(A_{G_k}). \tag{6}$$

This function corresponds to a probability of firing within a small time interval. Thus, it is expressed in Hz. Consequently, α_G is expressed in Hz mV^{-1} and N_G^{max} in Hz. A non-linearity is fixed so as to impose an upper limit over the firing rate. Here, it is modeled by a piece-wise linear function :

$$\mathcal{N}_G\left(V\right) = \left\{\begin{array}{ll} 0, & \text{if} \quad V \leq 0; \\ \alpha_G(V - \theta_G), & \text{if} \quad \theta_G \leq V \leq N_G^{max}/\alpha_G + \theta_G; \\ N_G^{max}, & \text{else.} \end{array}\right. \tag{7}$$

We have, for the RGCs gain control:

$$\mathcal{G}_G(A) = \left\{\begin{array}{ll} 0, & \text{if} \quad A \leq 0; \\ \frac{1}{1+A}, & \text{else.} \end{array}\right. \tag{8}$$

which actually differs from the non-linearity in the BCs gain control, following (Chen et al., 2013). The gain control rate, h_B for BCs, h_G for RGCs, expressed in Hz mV^{-1}, tunes the intensity of the gain control. In particular, if $h_B = 0$, $A_{B_i} \to 0$ exponentially fast so that the gain $\mathcal{G}_B(A) = 1$. The same remark holds for RGCs.

Parameters values of the model can be found in the appendix 5.1.

2.2 The cortical model

This is a two dimensional model composed of two populations of cortical columns: excitatory (E) and inhibitory (I), located in a cortical area of dimension $\alpha L_x \times \alpha L_y$, where α is a magnification factor from retina to cortex. Cortical columns represent the spatial average of cortical neurons at a space scale roughly corresponding to one pixel of voltage sensitive dye imaging (VSDI signal) (Chemla, 2010). This model has actually been employed in (Chemla et al., 2007) to reproduce the VSDI response to a simple visual stimuli (apparent motion) in the awake monkey primary visual cortex (V1). Cortical spatial coordinates are noted $(x, y) \equiv \vec{x}$. We thus use the same notations as for the retina, to alleviate notations, although there is a magnification factor between these two systems of coordinates.

The activity of cortical columns is represented by their average firing rate: ν_E for excitatory columns, ν_I for inhibitory columns. The equations for cortical neurons dynamics are based on a mean-field model of Adapting Exponential (AdEx) neurons (DiVolo et al., 2019). This model was derived under the hypothesis that the network dynamics is Markovian at a timescale of a few ms and stationary for a duration T. One describes then the collective dynamics through a master equation formalism developed by El Boustani and Destexhe (ElBoustani and Destexhe, 2009). This system can reproduce asynchronous irregular regime, a typical feature of the awake states, as well as Up and Down states, characteristic of sleep or anesthesia states (DiVolo et al., 2019).

The spatially extended dynamical system reads:

$$\begin{cases} T\,\frac{\partial \nu_E(\vec{x},t)}{dt} = -\nu_E(\vec{x},t) \;+\; F_E\left[\,\nu^{aff}(\vec{x},t) \;+\; \nu^{drive} \;+\; A_E^E\,\nu_E^{input}(\vec{x},t)\,,\; A_E^I\,\nu_I^{input}(\vec{x},t)\,\right] \\ T\,\frac{\partial \nu_I(\vec{x},t)}{dt} = -\nu_I(\vec{x},t) \;+\; F_I\left[\,\nu^{aff}(\vec{x},t) \;+\; \nu^{drive} \;+\; A_I^E\,\nu_E^{input}(\vec{x},t)\,,\; A_I^I\,\nu_I^{input}(\vec{x},t)\,\right], \end{cases} \tag{9}$$

where $\nu_E(\vec{x}, t)$ (resp. $\nu_I(\vec{x}, t)$) is the population rate of the excitatory (resp. inhibitory) cortical column located at $\vec{x}$, at time t. T is the characteristic integration time.

In eq. (9) the functions F_E (resp. F_I) are the transfer functions of excitatory (resp. inhibitory) neurons. They describe the firing rate of population E (resp. I) as a function of the excitatory and inhibitory rates ν_E and ν_I. Their form is made explicit in the appendix, section 5.2. The term $\nu^{aff}(\vec{x}, t)$ in eq. (9) corresponds to the retino-thalamic input (sensory drive). As the thalamus is not considered here (we assimilate it to a simple relay) this input comes directly from the RGCs. There is a direct correspondence, a retinotopy, between a point in the retina (RGC), and a point in V1 (cortical column). Here, this mapping is linear. There is just a magnification factor α from the retina to V1. Each RGC inputs a cortical column and $\nu^{aff}(\vec{x}, t)$ is the firing rate emitted by the RGC that inputs the cortical column located at $\vec{x}$. Note that, in the absence of gain control or

amacrine connectivity, the retinal model is reduced to a convolution cascade (Souihel and Cessac, 2021; Cessac, 2022; Kartsaki et al., 2024). In this case, the input $\nu^{aff}(\vec{x},t)$ is therefore similar to the one used in (Benvenuti et al., 2020). In the transfer functions F_E, F_I, we include a spatially uniform external drive ν^{drive}. This drive represents the background constant input coming from the rest of the brain.

The $\nu_E^{input}(\vec{x},t)$ (resp. $\nu_I^{input}(\vec{x},t)$) are excitatory (resp. inhibitory) inputs coming from the column itself or from other cortical columns. They are multiplied by an amplification connectivity factor A^{pre}_{post}, where pre and $post \in E,I$ stands respectively for "pre synaptic" and "post synaptic". In our case $A^E_E = A^I_I = A^I_E = 1$ and $A^E_I = 1.5$. This connectivity corresponds to the observed physiology of the real cortex. We have:

$$\left\{\begin{array}{l} \nu_E^{input}(\vec{x},t) \;=\; \int_{\mathbb{R}} \mathcal{N}_E(\vec{x}-\vec{x}')\,\nu_E(\vec{x}',t-||\vec{x}'-\vec{x}||/v_c)\,d\vec{x}' \\ \nu_I^{input}(\vec{x},t) \;=\; \int_{\mathbb{R}} \mathcal{N}_I(\vec{x}-\vec{x}')\,\nu_I(\vec{x}',t-||\vec{x}'-\vec{x}||/v_c)\,d\vec{x}' \end{array}\right. \tag{10}$$

where $\mathcal{N}_E, \mathcal{N}_I$ are 2D circular Gaussian connectivity kernels with mean-square deviation σ_E, σ_I:

$$\mathcal{N}_X(\vec{x}) \;=\; \frac{e^{-\frac{1}{2}\frac{\|\vec{x}\|^2}{\sigma_X^2}}}{2\pi\sigma_X^2}, \tag{11}$$

with $X = E, I$. This form follows e.g. (Zerlaut et al., 2018), although our connectivity kernel is two dimensional in contrast to their paper. Thus, σ_E, σ_I control the two dimensional cortical extent of the excitatory and inhibitory connections. Note, that, due to the normalisation of the Gaussian the shorter the cortical extent the larger the amplitude of the synaptic weight $\mathcal{N}_X(\vec{x})$ (see section 3.2.3 for a consequence of this). This normalisation means that the number of connections within the system remains constant when the size of the extensions is increased. The parameter v_c is the speed of axonal conduction (assumed to be a constant). Equations (10) therefore express that the excitatory input $\nu_E^{input}(\vec{x},t)$ is the sum of the incoming excitatory activity from the connected columns arising with a delay $||\vec{x}'-\vec{x}||/v_c$ depending on the distance between the columns and the axonal conduction speed.

The cortical model is designed to reproduce the pixels intensity of voltage sensitive dye imaging, corresponding to a variation of fluorescent luminosity with respect of the fluorescent baseline. Its expression depends on the average membrane potential of the excitatory and inhibitory populations in the column located at $\vec{x}$, at time t. Its equation is given by (22) in the appendix, section 5.2.

Remark. Note that this type of cortical model can exhibit pathological (i.e. model induced) oscillations due to bifurcations when some parameters become too large (such as ν^{drive} or the conduction speed). This phenomenon is well known and has been reported in the literature (Veltz, 2011). We observe as well such oscillations when parameters such as the retinal output amplitude becomes too large. This is commented in the afferent section.

2.3 Specific features of the model

In addition to propose an original implementation of the retino-cortical V1 system, the model has specific features. First, this is a two dimensional model in space, allowing to play complex trajectories and actually, to simulate the response to realistic visual scenes. This is a feature of the simulator we use, called Macular, developed at Inria. Macular is a simulation platform for the retina and the primary visual cortex (V1), designed to reproduce the response to visual stimuli, in normal vision conditions, or in altered conditions (pharmacology, pathology, development). It is organised into a layered structure that mimics the multi-layers organisation of the retina. It is fed by visual inputs (movies) then processed by the multi-layer structure. A description can be found at `https://macular.gitlabpages.inria.fr/macular/user_doc/Macular/main.html`.

In contrast to previous models which were considering retino-thalamic entries as spatially periodic functions (rings), the mere fact of introducing realistic retinal inputs leads to a distinction between left and right, bottom and top so that one cannot consider rings geometries anymore. In our case, we need to define proper boundary conditions to the model. Here, we assume zero boundaries conditions for the retinal and the cortical model. In contrast to ring topologies, this induces spatial inhomogeneities, even when considering the rest state.

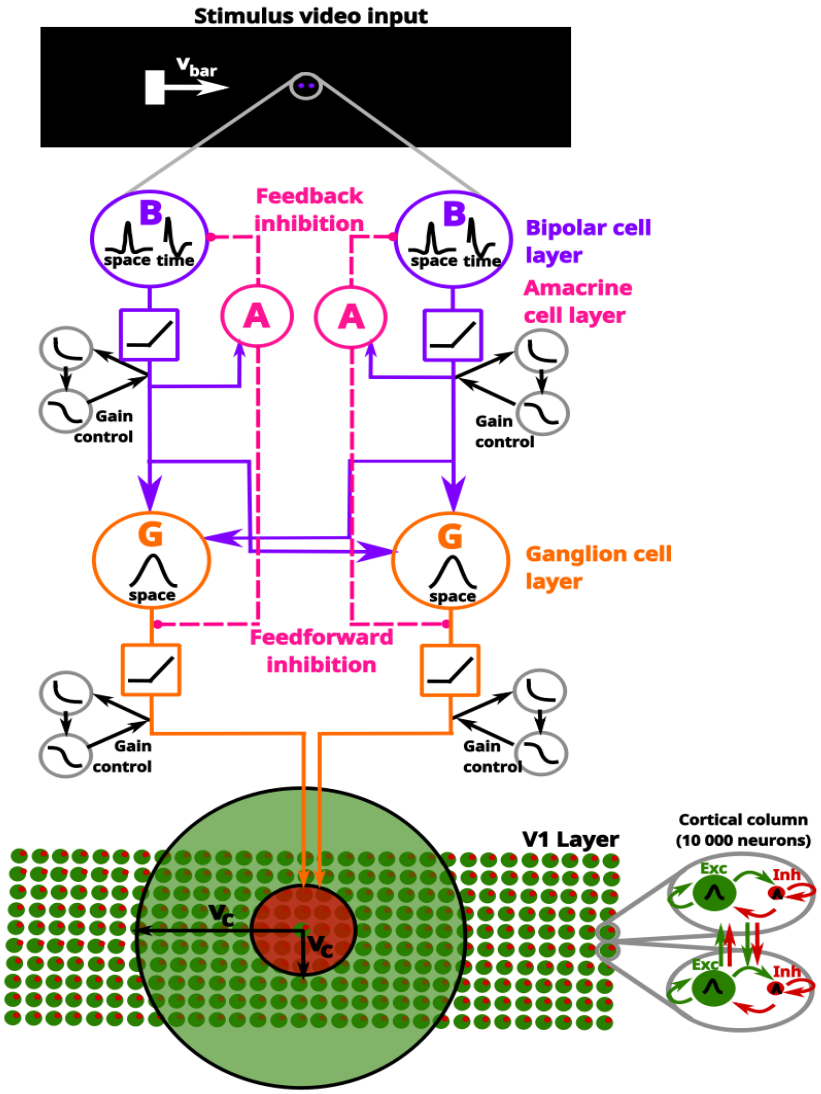

Figure 1: **Synthetic view of the retino-cortical model.** A stimulus is perceived by the retina, triggering a response. **From top to bottom**: The stimulus is first convolved with a spatio-temporal receptive field (black traces labelled "space" and "time" in the purple circles), that mimics in the Outer Plexiform Layer (OPL) the concerted activity of photoreceptors and Horizontal cells, and is fed to Bipolar cells (purple circles). This response is rectified by a low voltage threshold (purple squares). Bipolar cells responses are then pooled to retinal Ganglion cells (orange). The firing rate response of a Ganglion cell is a sigmoidal function of the voltage (orange square). Gain control can be applied at the Bipolar and Ganglion cells level (grey circles) triggering anticipation by a shift in the time to peak. The Bipolar cells activity is modulated by lateral inhibition through Amacrine cells (pink). The Ganglion cells response (firing rate) is sent to cortical columns in the primary visual cortex depicted, at the bottom right, as two interconnected mean field units (small circles) corresponding respectively to excitatory (green) and inhibitory (red) population. A sketch of this interconnection is shown on the right zoom (grey ellipses). Cortical columns are connected together by an excitatory (big green circle) and inhibitory (big red circle) horizontal connectivity. Note that we assume the same conduction velocity v_c for both connectivities.

Finally, why is this model a "chimera" ? While the design and parameters of the cortical model are based on previous works on the monkey visual cortex, the retinal model is essentially based on previous works dealing with mice retinas. In addition, as mentioned above, there is no thalamus, or, more precisely, it is transparent, considered as a simple relay. In this sense, our model resembles more a "Frankenstein" creature than a real organism. However, we believe that it captures the main mechanisms in the retino-cortical entanglement. Also, the advantage of such a model, with its joint simulation platform, is that one can easily modify the parameters of such or such components and see how it modifies the observed response. This is actually the main philosophy of this work where we vary physiological parameters, such as the conduction speed, or the effect of amacrine cells, that cannot be easily varied experimentally. This also means that the chimera model could be made closer to monkeys by adjusting the retina model to experimental results on monkeys retinas. Note that a feature of the simulator Macular is precisely to afford such changes.

2.4 Observables for anticipation

From now, both simulated retinal and cortical areas are 2D spaces of 81×15 retinal cells. As two consecutive retinal or cortical columns are spaced by $0.225°$ of visual angle these areas correspond to $18.225° \times 3.375°$. Using the conversion factor of about 0.3 mm per degree in the retina, approximated from `https://www.ncbi.nlm.nih.gov/books/NBK11556/` and 3 mm per degree in the cortex for humans this gives a retinal area of $\sim 5.46 \times 1\ mm^2$ and a cortical area of $54.6 \times 10\ mm^2$ with a spacing of $67.5\ \mu m$ between retinal cells and $675\ \mu m$ between cortical columns. We discard the first and the last horizontal degree in all figures to reduce boundaries effects, giving thus an effective cortical space of $16.45°$ long.

In this paper, we mainly consider the motion of a bar, moving horizontally, from left to right, along the x axis with a constant speed v_B. The bar starts to move at time $t = 0$ where the center of the RF of the cortical column located at $x = 0$ coincides with the middle of the bar. The bar has dimensions 0.67×0.9 degrees of visual angle, i.e. its height is small compared to the vertical extent of the retina and the cortex. As a consequence, we will consider, in the definition of anticipation observables, that the bar response is characterized by two dimensional graphs, with one spatial dimension, x, and the time, t.

2.4.1 Retinal anticipation

Anticipation has been observed in the retina, according to different modalities. The first characterization was provided Berry et al. (Berry et al., 1999). It is characterized by a temporal shift in the peak of the ganglion cell response to a moving object, occurring before the peak response to the same object when flashed (Berry et al., 1999; Chen et al., 2013). This can be explained by gain control which has the effect of advancing the peak response of the cell's activity. In our model, this effect can arise at the level of BCs or RGCs. A detailed study was published in (Souihel and Cessac, 2021). It shows that, with gain control, anticipation time grows with the size and the contrast of the bar while it decreases with its velocity. In our retina model, amacrine cells can also induce

anticipation by advancement of the peak, independently of gain control, although these two effects can constructively combine. We qualify this peak shift mechanism as *peak anticipation* or *adapting anticipation* (using the terminology of (Johnston and Lagnado, 2015)) since BCs and RGCs adapt according to their level of activity. Therefore, we quantify retinal anticipation by a RGC peak shift and we define the two peak-based quantities: the time at which the peak of the response occurs (Time To Peak, TTP); and the delay between the arrival of the bar at the centre of the receptive field and the time of the peak (Stationary Peak Delay-SPD). These quantities are defined in detail in the next section, as they are also used to characterize cortical anticipation.

2.4.2 Cortical anticipation

Anticipation was also observed in the primary visual cortex (V1). In this section, we define several observables related to the cortical activity when responding to the moving bar. These observables are here introduced in a case where the sensory drive of the cortical model consists of the retinal response to a white bar with neither gain control nor amacrine cells. Thus, the retina is passive. We refer this as control conditions (CTL). All parameters are given in the table 1 of the appendix. The illustrative figure 2 corresponds to a bar moving at $6°/s$ (equivalent to 18 mm/s in the cortex). This is also the value of the default bar speed.

The indicators for cortical anticipation are based on the typical VSDI signal curves, shown in Fig. 2A, and reproducing the experimental observations made by (Benvenuti et al., 2020). We first define quantities attached to individual columns, i.e. depending on the spatial coordinate x.

Local observables.

- $t_{center}(x)$ is the time when the middle of the bar reaches the center of the receptive field of the cortical column located at x. We have thus $t_{center}(0) = 0$.
- **The activation time (AT)**, $t_{ON}(x)$, is the time when the VSDI signal response becomes larger than a threshold $\theta = 0.001$. Note that we also tested a criterion based on a threshold on the derivative of the VSDI signal. However, it was not reliable, requiring ad hoc smoothing and leading to spurious artefacts.
- **The latency** is $t_L(x) = t_{ON}(x) - t_{center}(x)$. Since $t_{ON}(x) \leq t_{center}(x)$, the latency can be negative or equal to 0. The latency increases therefore when its value becomes more negative.
- **The time to peak (TTP)** is the time $t_P(x)$ when the VSDI signal reaches its maximum (see Fig. 2B).
- **The peak delay** is $t_{PD}(x) = t_P(x) - t_{center}(x)$. This quantity can, a priori, have any sign, although in CTL conditions it is always positive. This value can be negative though due to e.g. to the retina influence.

The latency, $t_L(x)$, depends on the distance of the RF of the cortical column to the position where the bar started. To illustrate this dependence we use a color code,

similarly to the paper (Benvenuti et al., 2020). The different RF positions in the visual field are represented by a color gradient (see Fig. 2B), from black ($x = 0$ degree, point where the bar starts) to light blue ($x = 18.45$ degree). Although, for a spatio-temporal representation x, t, there is a redundancy between the x coordinate and the color, this representation is actually kept for didactic purposes. Especially, as shown in Fig. 2B, when plotting all cortical responses for various trajectory lengths, aligned on $t_{center}(x)$ (so that the time variable is $t - t_{center}(x)$), one observes a change in the shape of the response prior to the peak: the latency increases with the distance to the bar origin, i.e. as the distance of the cortical columns to the bar origin increases, these columns respond earlier and earlier to the bar propagation. Following (Benvenuti et al., 2020) we interpret this as a *cortical anticipation by latency*. We now introduce global observables to quantify this form of anticipation.

Global spatial observables. Although the latency corresponds to a time as a function of space, it is useful to invert the axes and, instead, to represent space as a function of time: the curve $t_L(x)$ becomes $x(t_L)$ by a simple symmetry with respect to the first diagonal. The cortical space, ploted in function of the activation time and the time to peak have characteristic shapes illustrated in Fig. 2C. This allows us to define four spatial observable, following Benvenuti et al (Benvenuti et al., 2020):

- **The anticipation range (AR)** is the maximal spatial limit of anticipation, the maximal distance at which cortical columns interact to anticipate motion, as argued in (Benvenuti et al., 2020). It corresponds to the ordinate of the inflexion point in Fig. 2C. Beyond this spatial position, the working hypothesis is that anticipation by latency saturates, due to the limited cortical extent of horizontal connectivity which causes anticipation (Angelucci et al., 2002). Consequently, beyond the anticipation range, the slope of the activation time becomes constant (Fig. 2C), the activation time is no longer affected by the horizontal cortical propagation, but only by the bar speed.

- **The short-range activation speed (SRAS)** is the slope of the activation time before the inflection point (Fig. 2 C). When plotted as a function of time from stimulus onset, this speed, actually sums up two different contributions: the bar speed and the anticipation speed carried by horizontal connectivity (Fig. 2C). We took care to subtract the bar speed so that the SRAS corresponds only to the contribution of lateral (intra cortical) connectivity.

- **The Long range activation speed (LRAS)** is the slope of the activation time after the inflection point. This cortical region is characterized by an absence of horizontal cortical connectivity influence. In our model, its slope is only constrained by the bar speed. We mention this quantity because it is expected to vary due to feedback effects, not present in our model (Benvenuti et al., 2020).

- **The Peak speed (PS)** is the slope of the time to peak curve. It is parallel to the moving bar curve.

Global temporal observables. One can also plot x as a function of the latency (see figure 2D). We obtain two temporal observables, illustrated in Fig. 2D:

- **The maximal latency (ML)**. One observes that the latency of cortical columns close from the starting point of the bar (black circles) increases until it saturates, as the columns are located further and further from the starting point of the bar. The asymptotic value is the maximal latency.
- **The Stationary Peak Delay (SPD)**. In contrast to ML, the peak delay is actually independent of x, in all the cases studied in the paper. We discuss this homogeneity in the discussion section. We call this value the SPD.

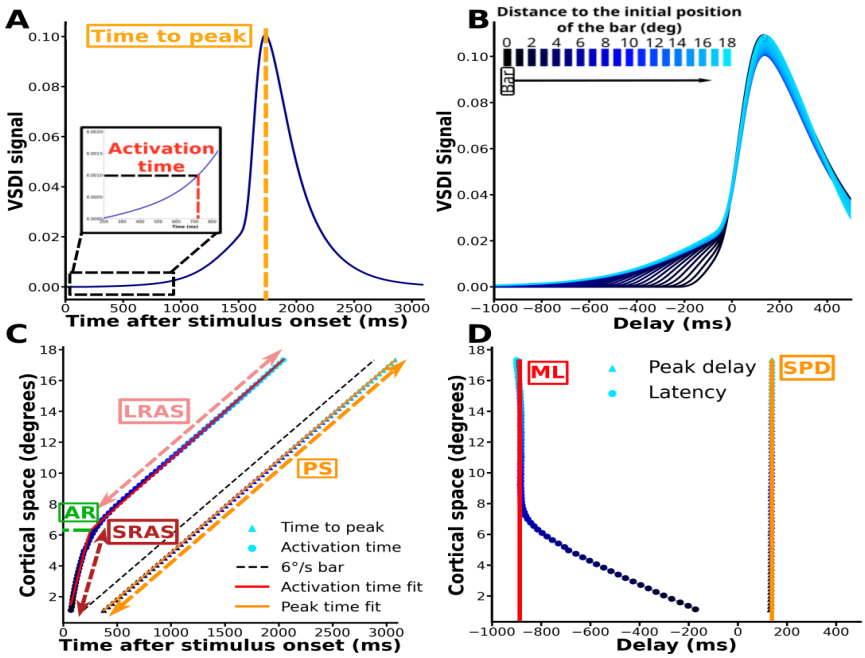

Figure 2: **The paradigm of cortical anticipation by latency. A) Typical time course of the VSDI signal response** for a fixed cortical column located at x. The time to peak corresponds to the yellow dotted line. The insert shows a zoom of the proximal part of the curve where the activation time is computed (the black dotted line corresponds to the threshold θ, see text.). **B) Global representation of the spatio-temporal VSDI responses using a collapse of the VSDI signals**. The curves corresponding to cortical columns located at different spatial locations are aligned on $t_{center}(x)$. Then, these curves are colored so that the different receptive fields positions in the visual field are represented by a color gradient (displayed on the top of the figure), from black ($x = 0$ degree, point where the bar starts) to light blue ($x = 18.45$ degree). The legend at the top left represents the color gradient (from black to light blue) associated with the position of the cortical column. **C) inverted time space representation** where the cortical space (x coordinate, in degrees of visual angle) is represented as a function of time (in ms). The dotted black line corresponds to the displacement of the center of the moving bar. This is a straight line, $x = v_B\, t$ (the center of the bar is located at $x = 0$ when $t = 0$). The *time to peak*, $t_P(x)$, is represented by triangles, colored according to the color gradient. The curve (orange line) is a straight line too, with equation $x = v_B\,(t - t_P(0))$, where $t_P(0)$ is the time to peak for the column located at $x = 0$. The *peak speed* (PS) is the slope of this curve (thus, in CTL conditions, this is the bar speed, v_B). The activation time, $t_{ON}(x)$ is represented by circles colored according to the color gradient. The curve represents a crossover between two regimes, well fitted by straight lines (red curves), and separated by an inflection point. The spatial position of this inflection point is the *anticipation range (AR)*. The slope of the first linear part (dark red dashed arrow) corresponds to the *short-range activation speed (SRAS)*. The slope of the second linear part (light red dashed arrow) is the *long range activation speed (LRAS)*. In CTL conditions, this slope is equal to v_B. **D) Maximal latency (ML) and Stationary Peak Delay (SPD)**. The cortical space is represented as a function of the latency (in ms). The latency increases until saturation (red vertical line), at a time called *maximal latency (ML)*. The peak delay (triangle) is constant and equal to the Stationary Peak Delay (SPD), (yellow vertical line).

Interpretation. The figure 2 essentially shows that the moving bar triggers intricate waves of cortical activity, transmitted by the columns sensing the bar to the distant columns via the horizontal connectivity, at a time influenced by both the bar speed and the axonal velocity (Fig. 2B and C). In other words, cortical columns sense the bar earlier than the time when the bar enters their receptive field. This reproduces the observations of Benvenuti et al. (Benvenuti et al., 2020) and was interpreted by the authors as an indication that cortical anticipation mainly holds by latency. This is the explanation of the change in the early VSDI response profile (before the peak), in Fig. 2B, as a function of the color gradient. One observes therefore that the latency becomes more and more negative with the distance of the cortical column to the position where the bar starts, Fig. 2D, meaning that the cortical columns are informed earlier and earlier that something is arriving. This observation holds until a saturation value, the maximal latency, where the gain of anticipation provided by the horizontal cortical connectivity reaches a maximum. This suggests that cortical connectivity and axonal velocity impacts anticipation (sections 3.2.3, 3.2.4). As developed below, we also observe an anticipation by peak shift, where the peak in the VSDI signal is advanced with respect to the peak in the RGCs response.

3 Results

3.1 Model calibration

The model was calibrated so as to reproduce the results obtained by (Benvenuti et al., 2020) in the monkey primary visual cortex. Note however that, although our color gradient is similar to the one used by these authors, it corresponds to different positions. Indeed, in our simulations, the cortical area and the visual field used are bigger than in their experiments (resp. 54 mm versus 21 mm, and 18 degrees versus 7 degrees). A larger simulated cortical area was indeed needed to be able to correctly observe the effect of certain parameters such as bar speed or cortical extension.

The figure 2 illustrates the result of our model in CTL conditions, with a bar speed of $6^\circ/s$. It reproduces the shape of the latency curve shown in (Benvenuti et al., 2020), Fig. 4: cortical columns far from the bar origin are activated earlier than the ones close to the bar origin. This curve is divided in two regimes commented in section 2.4 and illustrated in Fig. 2. In the first regime the slope (short-range activation speed) have a value of $24.8^\circ/s$ in experimental data and of $21.3^\circ/s$ in simulated data. The inflection point (anticipation range) differs though: 2° in experiments and 6.3° in simulations. Also, the activation in our model can occur up to 880 ms before the middle of the bar arrives at the center of the receptive field (i.e. a maximal latency of -880 ms) while in (Benvenuti et al., 2020) it is said that "latency scatter for the medium and long trajectories that fully covered a wide range of values from 0 to -400 and -800 ms". We finally note a difference in the time of the peak, very close to zero in experiments (Fig. 4D of (Benvenuti et al., 2020)) but located at 139 ms in the simulation. To understand these discrepancies, it is important to note that our simulations are made in control conditions i.e. without any retinal anticipation mechanism, whereas such mechanisms are presumably present in experimental conditions. Indeed, as commented later, adding

retinal anticipation reduces these differences (section 3.3). Taken together, these results provide a good basis for further explorations on the role of retinal and cortical effects on anticipation with discrepancies expected to be reduced in the presence of a realistic retinal input.

3.2 Cortical anticipation depends on stimulus features and on physiological parameters

We now study, in CTL conditions, the dependence of anticipation on several parameters, such as the bar speed or contrast, and physiological parameters, such as the conduction speed following the modalities described in section 2.4.2. The default value of the parameters, including the bar speed, are those reported in the Appendix 5.1.

3.2.1 Increasing the retinal output amplitude enhances anticipation

We varied the retinal output amplitude (ROA), sent by retinal ganglion cells, in a range from 1 to 50 Hz (figure 3). For this, we increased the amplitude $\mathcal{C}$ of the OPL kernel (eq. (1)). This has the effect of increasing the firing rate of the RGCs in a linear way. Note that we are still in CTL conditions here, the retina is passive. Increasing the ROA increases almost linearly the VSDI signal (Fig. 3 E) with a slight saturation presumably due to the sigmoid activation functions F_E, F_I in the mean field model of cortical columns (eq. (9)). Here, we chose to stay in the almost linear range. Indeed, we observed that increasing too much the amplitude of the input leads to pathological oscillations which, as mentioned above, are artefacts of the cortical model.

The peaks of VSDI signal, shown in Fig. 3 A (1 Hz) and Fig. 3 B (35 Hz), has globally the same shape as the paradigmatic figure 2A, although we observe a clear difference in the latencies between A and B. This is a first indication that increasing the ROA enhances anticipation by latency. Fig. 3 C shows the time course of the VSDI signal for the cortical column located at the center of the layer ($x = 9°, y = 1.35°$). One observes an increase in the slope before the peak, resulting in an increase of the maximal latency (ML) which becomes more negative as shown in Fig. 3 F (red trace). There is a saturation for large ROA though. Along the same lines, we plot, in 3 D, the spatial VSDI signal for the time when the central cortical column reaches its maximum. Note that the x coordinates has been shifted so that the central cortical column is actually located at $x = 0$ in this figure. We observe a spread of the left part of the peak and a more abrupt slope on the right part, as the ROA increases. Those combined effects results in a increase of the anticipation range (AR) as shown in Fig. 3 F (green trace), with again a saturation effect. Finally, the spatial and temporal effects combine to increase the short range anticipation speed (SRAS). Therefore, anticipation by latency becomes more prominent (ML and AR) and spreads faster (SRAS) as the ROA increases.

Interestingly, we also observe a slight anticipation by peak shift in Fig. 3 C. For a more quantitative study we have plotted, in Fig. 3 H, the SPD as a function of the ROA. It is constant for RGCs, but essentially decreases for the cortex. This figure also shows the difference between the cortical and retinal SPD, which is negative, meaning that the VSDI peak arises earlier than the RGC frequency peak. Thus, in addition to show

anticipation by latency, the VSDI signal is also a bit in advance on the RGC peak. It is enhanced when the ROA increases, up to some maximum at about 35 Hz, where the difference is maximal, i.e. where the VSDI peak anticipates the most the RGC frequency peak. However, beyond 35 Hz, we start to observe (small) oscillations in the VSDI signal so that the increase of the curve after 35 Hz might be an artefact of the model. It is interesting to note that even at small retinal amplitudes (1Hz), the cortex remains 8.8 ms ahead of the retina. In Fig. 3 I we have plotted the time course of the frequency response near the peak for the central RGC, and, in Fig. 3 J, the time course of the VSDI, near the peak for the central cortical column. The red traces correspond to a 1 Hz ROA and the green traces to 35 Hz. The green traces have actually been rescaled to match the amplitude of the red ones. This is to show that, in addition to a simple rescaling (which makes the RGC response overlap in Fig. 3 I), there are, in the cortex, non linear effects which modify the shape of the response and thereby impact the anticipation.

To summarize, an increase in the ROA non-linearly enhances the anticipation which extends further, earlier and faster, with a saturation when the ROA becomes too large. Rising the ROA increases the overall activity level transmitted horizontally by cortical columns. Thus, more distant cortical columns are above the threshold earlier as shown by the increase in ML and AR. The results also demonstrate the presence of a mechanism in the cortex enabling anticipation by peak shift, an effect which increases with the ROA. However, while in anticipation by latency the shape of the response curve depends on the distance of the cortical column to the starting point of the bar motion, the peak shift is independent of this.

Stimulus contrast. We assessed the impact of the stimulus contrast in the movies by increasing it from 0.1 to 1 by increments of 0.1. As we checked (Fig. 4) the effect is completely equivalent to increasing the amplitude $\mathcal{C}$ of the OPL kernel and thus, the ROA.

3.2.2 Cortical anticipation non monotonously depends on the bar speed

We have done simulations with stimulus speeds ranging from $3\,^\circ/s$ (equivalent to 9 mm/s of the cortex) to $30\,^\circ/s$ (resp. 90 mm/s in the cortex), still in conditions where the retina is passive, with the default parameters of Appendix 5.1. Our results are summarized in Fig. 5. The first remark is that, increasing the bar speed decreases the retinal output amplitude (ROA). This is because the OPL (outer plexiform layer) kernel $\mathcal{K}_{B_i}$ in eq. (1) has less time to integrate the stimulus. We actually see the decrease in the BCs activity and RGCs activity as the bar speed increases (Fig. 5 A). This induces a decrease in the VSDI signal activity (Fig. 5 B). Thus, from the conclusions of the previous section, one expects a reduction of the anticipation by latency. Increasing the bar speed indeed diminishes the maximal latency (ML) and the anticipation range (AR) (Fig. 5C).

Is this effect on ML and AR only due to the ROA reduction or are there more subtle, non linear, effects hidden ? To address this question we plotted ML, AR as a function of the ROA itself controlled by the bar speed and compared to the case of Fig. 3 F where the ROA was under direct control. (Fig. 5 D). For the range of bar speeds that we explored the ROA varies in the interval $[10, 45]$ Hz, a bit less that the range of ROA explored in the previous section. In the common interval of variation, we observe that the AR,

when it is controlled by a direct variation of the ROA (dark green curve) behaves almost linearly, similarly to the case when the ROA is tuned by the bar speed (light green curve), although with a smaller slope in this case. In contrast, the ML increases non linearly, and goes to a saturation when the ROA is directly modified (red trace) whereas it increases sharply and slightly non linearly in the case where the ROA is modulated by the bar speed (brown trace). This evidences that the decay in ML and AR versus the bar speed is not only due to the decay in the ROA but includes additional, non linear effects.

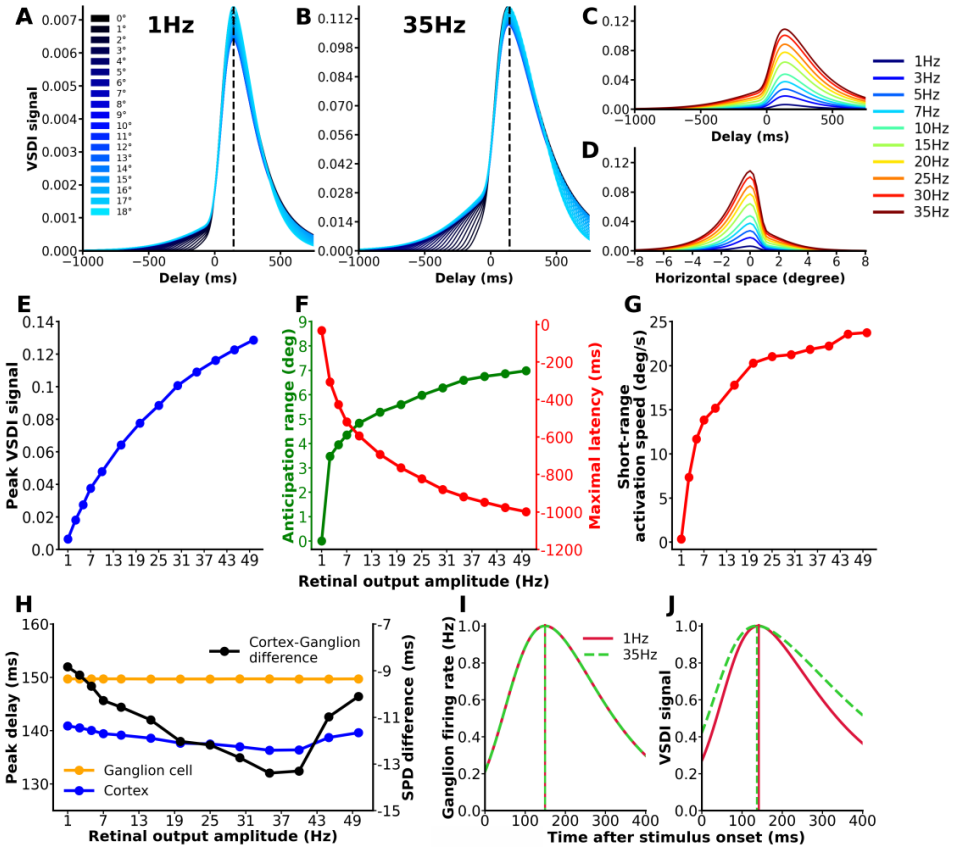

Figure 3: **The effect of the retinal output amplitude (ROA) on the cortical response. VSDI signal response to ROA at A)** 1 **Hz and B)** 35 **Hz.** The color bar on the left of Fig. A represents the color gradient introduced in section 2.4. **C) Temporal VSDI signal** in response to increasing retinal amplitude for the cortical column located at the center of the layer ($x = 9°, y = 1.35°$). **D) Spatial VSDI signal** in response to increasing retinal amplitude, for the time where the central cortical column reaches its maximum. The x coordinates has been shifted so that the central cortical column is actually located at $x = 0$. **E) VSDI signal amplitude** of the central cortical column versus the ROA. **F) Temporal and spatial observables**: maximal latency (red, scale on the right) and anticipation range (green, scale on the left) versus the ROA. **G) Speed observable**: short-range activation speed (red) versus the ROA. **H) Retino-cortical stationary peak delay (SPD) variation** versus the ROA for the central cell. The ganglion firing rate SPD is plotted in orange and the VSDI signal in blue (scale on the left). In black, is represented the difference between the cortical and RGC SPD (scale on the right). A negative value means a cortical peak arising earlier than the RGC peak. **I) Shape of the central RGC response profile to the moving bar**, for 1 Hz (red) and 35 Hz (dashed green). Note that the red trace is normally quite smaller than the green trace, but we have rescaled it to show that the difference between 1 Hz and 35 Hz, at the retinal level, is only a rescaling. This contrasts with **J) VSDI signal**, where the same rescaling let also appear distortions due to the non linearities in the cortical model. The dotted vertical lines in I, J correspond to the peaks in the RGCs firing rate or VSDI signal at 1 Hz (red) and 35 Hz (green).

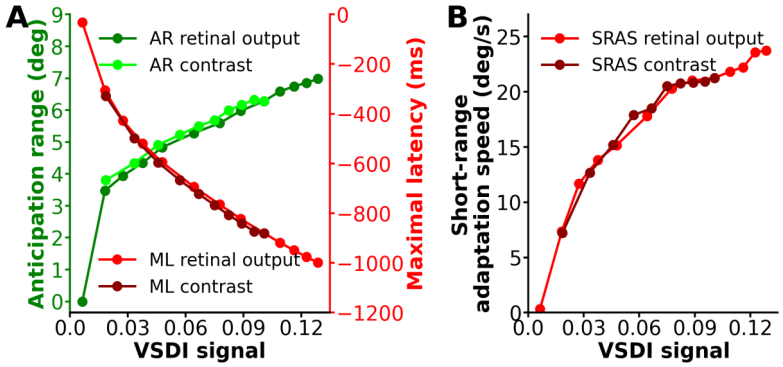

Figure 4: **Stimulus contrast and retinal output amplitude comparison. A) Temporal and spatial observables in function of VSDI amplitude**: Maximal latency with retinal output (red) or stimulus contrast (dark red) variant. Anticipation range with retinal output (green) or stimulus contrast (light green) variant. **B) Speed observable**: short-range activation speed with retinal output (red) or stimulus contrast (dark red) variant.

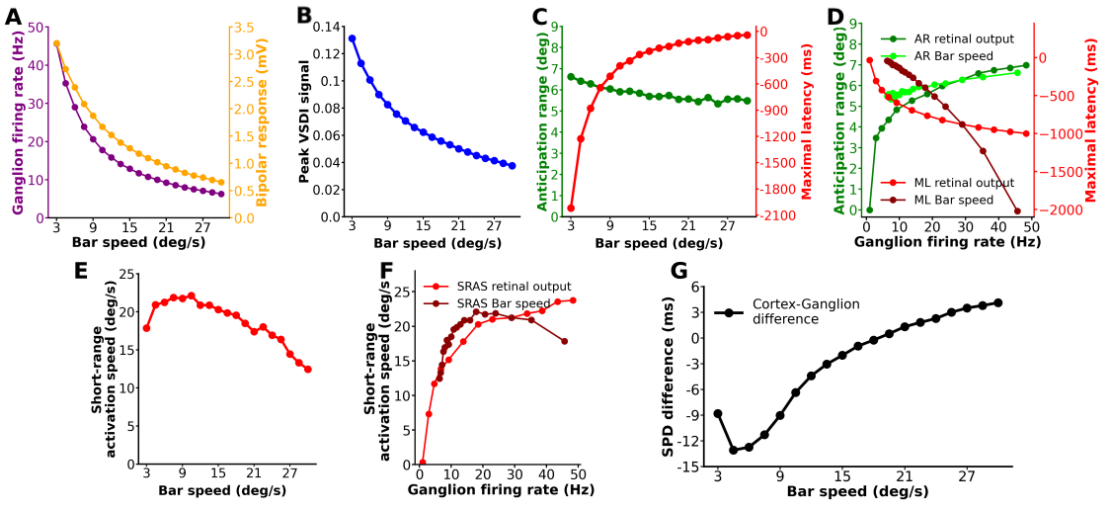

Figure 5: **The effect of the bar speed on the cortical response. A) decay** of the BC voltage peak and RGC firing rate (ROA) as the bar speed increases. **B) VSDI signal amplitude** of the central cortical column. **C) Temporal and spatial observables**: maximal latency (ML, red) and anticipation range (AR, green) in function of the bar speed. **D) ML and AR as a function of the RGC firing rate**, in the case where the firing rate is constrained by the ROA (red for ML, dark green for AR) and in the case where the firing rate is constrained by the bar speed (brown for ML, light green for AR). **E) Speed observable** : short-range activation speed (SRAS, red) versus the bar speed. **F) SRAS as a function of the ganglion firing rate**, in the case where the firing rate is constrained by the ROA (red) and in the case where the firing rate is constrained by the bar speed (brown). **G) Stationary peak delay difference** between cortical VSDI and ganglion firing rate.

Similarly, we studied the effect of the bar speed on the short range activation speed (SRAS), Fig. 5 E. This quantity shows an increase up to $9°/s$ (2.7 mm/s in the retina, 27 mm/s in the cortex), then a decrease, suggesting the existence of a range of preferred speeds where anticipation by latency is optimal. Note that this effect cannot be explained only by the decrease of the ROA, as shown by Fig. 5 F. The red trace (direct control of ROA) is rather different from the curve where ROA is controlled by the bar speed (brown trace).

One actually expects three distinct effects as the bar speed increases. First, a decrease of the VSDI signal amplitude, since the OPL convolution kernel has less time to integrate the stimulus, directly impacting the anticipation. Second, a cortical column integrates the retinal signal as well but if the speed of the bar is very small (say, even static) its response occurs within a characteristic time quite shorter than the time it takes to the bar to reach the next column. In other words, when the bar arrives to the next column, the activity coming from the previous one has dropped to zero. This does not allow the columns to build up a non linear propagating front travelling faster than the bar. As the bar speed increases this front takes place and anticipation gradually increases. However, the horizontal cortical connectivity has less time to build up long range excitation. Thus, as a third effect, the activity generated by this bar eventually catches up the one carried by the horizontal connectivity. In consequence, the speed of anticipation is gradually overwhelmed by the speed of the bar. This last effect explains the maximum observed in the SRAS. Beyond this point anticipation by latency is more and more driven by the decay of the VSDI signal, as the bar speed approaches the conduction speed (although we are far from this limit in our bar speed range).

Finally, we investigated the role of the bar speed on the anticipation by peak shift (Figs. 5 G). The stationary peak delay (SPD) shows up a minimum at about $4.1°/s$, where the advance of the cortical SPD with respect to the RGC SPD is maximal. After this minimum there is a non linear, sigmoidal like, increase of the SPD, which switches from negative to positive at about $16°/s$. Thus, for larger speed, the cortical peak is delayed with respect to the RGC peak. This effect of the SPD can be explained as follows. An increase in the speed causes a reduction in the RGC response (Fig. 5 A) and cortical SPDs before stabilising above $9\ °/s$.

To sum up increasing the bar speed first decreases the ROA. In parallel, one observes a monotonous decrease of ML and a (moderate) decay of AR. However, this detrimental effect on anticipation is not only due to the decay of the output; additional, non linear effect take place. This is prominent when observing the SRAS which shows a "preferred" speeds range (at about $9°/s$) where it is maximal. This preferred speed is also the place where the SPDs of RGCs and cortex saturate. We also observe a slight anticipation by peak shift (overwhelmed by anticipation by latency) with a "preferred" speed at about $4.1°/s$.

3.2.3 The role of excitatory and inhibitory connections length on cortical anticipation

We have next explored the influence of the excitatory and inhibitory connectivity lengths on cortical anticipation, an effect which cannot be studied experimentally (Fig. 6). We

need to recall first a salient feature of the model. The cortical connectivity is modeled by a *normalized* Gaussian kernel (section 2.2, eq. (11)) where the cortical extent (excitatory or inhibitory) is the mean square deviation of the Gaussian. This normalisation was introduced in the original model (Zerlaut et al., 2018) to ensure that the number of connections within the system remains constant when the size of the extensions is increased. As a consequence, the shorter the cortical extensions, the more the Gaussian connectivity profile is concentrated around the cortical column's receptive field, with a higher weight. Inversely, increasing the cortical extensions spreads the Gaussian and reduces its weights. Therefore, acting on the Gaussian mean square deviation dramatically influences the value of the mean cortical column voltage for the excitatory and inhibitory populations as well as their polarisation in the steady state, with, thereby a significant impact on the VSDI signal, anticipation range (AR), and maximal latency (ML). This is illustrated in Fig. 6.

We maintained the ratio between the two respective lengths to a constant ratio of 0.2 so as to keep the balance between excitation and inhibition extents. The excitatory extent was then varied from 1° to 7° and the inhibitory extent from 0.2° to 1.4°. The first prominent effect, observed in Fig. 6 A, is the behaviour of the VSDI signal amplitude. In contrast to the previous cases, it is non monotonous. It decays up to a minimum at about 2° before increasing. To better understand this behaviour we have plotted, in Fig. 6 F, the peak of the excitatory mean voltage (called $\mu_{V,E}$ in the appendix 5.2) as a function of the excitatory cortical extent (dark green) while the light green curve represents the excitatory mean voltage at steady-state. The difference between the two, "mean voltage difference", is shown in Fig. 6 H (dark green). Likewise, Fig. 6 G shows the peak of the inhibitory mean voltage (brown) and inhibitory mean voltage at steady-state (red), while the red trace in Fig. 6 H shows the difference between these two quantities. The total VSDI signal as a function of the cortical extent is a linear combination of these two traces, (eq. (22), appendix 5.2).

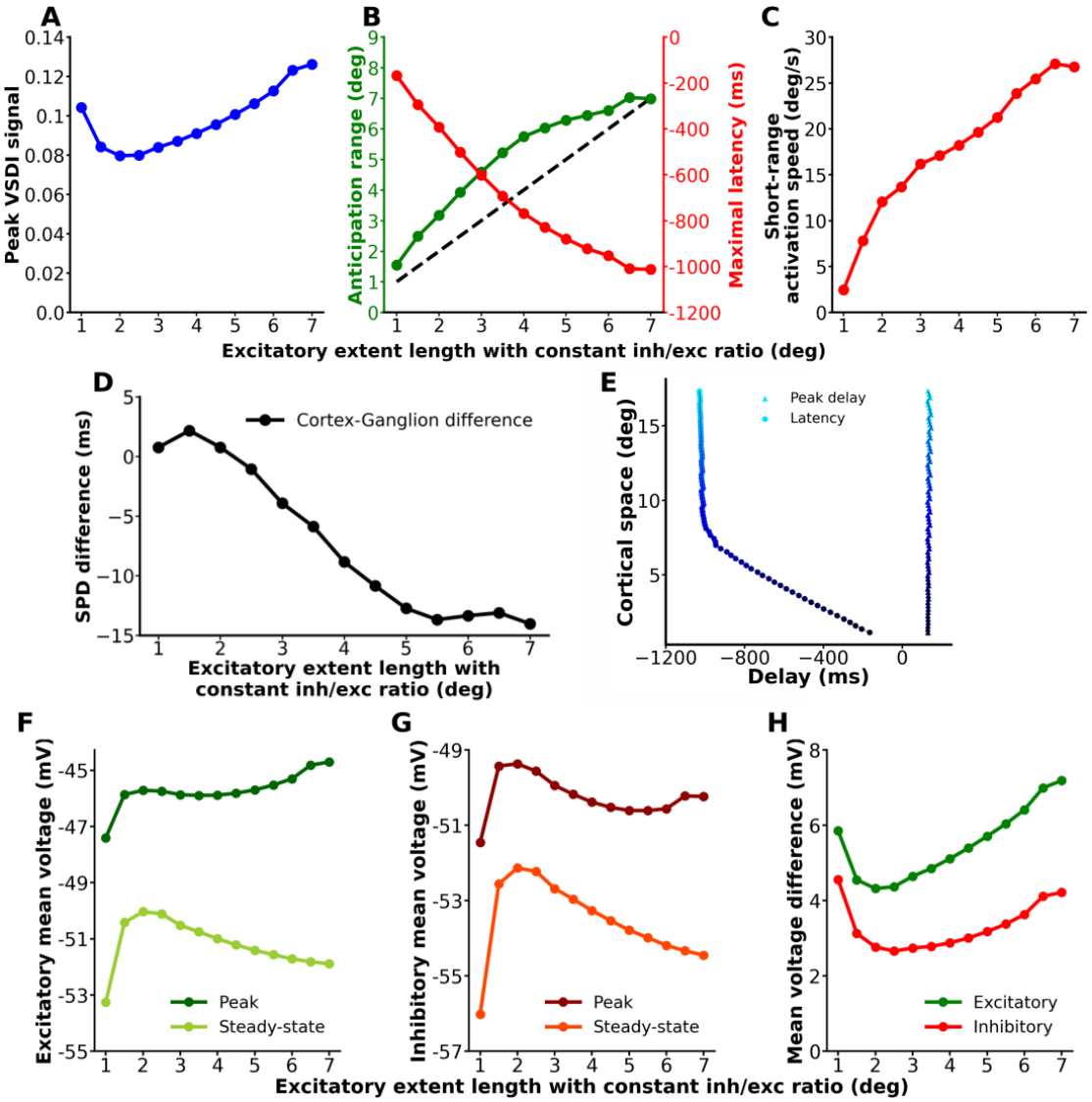

Figure 6: **The effect of excitatory and inhibitory connectivity length on the cortical response.** In the whole figure, the variable on abscissa is the excitatory cortical extent, but recall that there is a constant ratio with the inhibitory extent. **A) VSDI signal amplitude** of the central cortical column versus the cortical extent length. **B) Temporal and spatial observables:** anticipation range (green) and maximal latency (red) in function of excitatory. The black dotted line represents the equality between distance and length of excitatory connectivity. **C) Speed observable :** short-range activation speed (red) in function of excitatory extent. **D) Difference** between RGC stationary peak delay (SPD) and VSDI signal SPD. **E) Cortical space as a function of delay for excitatory extent at** 7° **and inhibitory extent at** 1.4°**.** Two delays are represented: the activation delay called latency (circle) and the peak delay (triangle) detailed in Sec. 2.4.2 and Fig. 2. **F) Excitatory mean voltage** at the peak (dark green) and at the steady-state (light green). **G) Inhibitory mean voltage** at the peak (dark red) and at the steady-state (light red). **H) Mean voltage peak amplitude** for excitators (green) or inhibitors (red). This is the difference between the average peak voltage and the steady state voltage.

When the cortical extent increases from 1° up to 2° we observe that the excitatory and inhibitory mean voltages increase while the mean voltage differences decreases, explaining the observed decay of the VSDI signal. This decrease is due to the fact that the mean voltage at steady-state increases faster than the peak mean voltage, for both population. As anticipated above, this behaviour can be explained by the Gaussian connectivity profiles. Increasing the cortical length decreases the intensity of the coupling between two cortical columns. Short extensions give a concentrated Gaussian with strong weights but short range influence. When the inhibitory cortical extent is 0.2° (resp. 1° for the excitatory cortical extent), the inhibitory Gaussian is so concentrated that its weight far exceeds that of the excitatory. This inhibitory dominance gives rise to a strong steady-state hyperpolarisation and an overall sensitisation of the cortical column which reacts more strongly to the stimulus. This hyperpolarisation is reduced by the elongation of the inhibitory extensions above 0.2°. This mean votage difference decreases leading to a reduction in the VSDI. From 2°, we observe a decay in the steady state voltage for excitatory and inhibitory population, corresponding to a new phase of hyperpolarisation at steady-state. When the cortical extent increases, the excitatory and inhibitory weights decrease thereby diminishing the mean voltage at steady-state. In contrast, the mean voltage peak increases for excitators and have a moderate variation for inhibitors. As a consequence, the mean voltage differences increase, leading to an increase of the VSDI.

We observe more anticipation by latency since AR and ML increase when the cortical extent increases (Fig. 6 B) as well as short range anticipation speed (SRAS) (Fig. 6 C) with a saturation for an excitatory and inhibitory length respectively at 7° and 1.4°. Here, it is interesting to note that the measured AR is always larger than the length of the corresponding excitatory connectivity (black dotted line, Fig. 6B). The cortical columns are therefore influenced beyond the excitatory extent, and, therefore also earlier, promoting anticipation. This is particularly true up to 4° where this effect is maximal. Beyond 7°, the AR and ML start to saturate. We believe that this arises because the Gaussian extent increases at the expense of proximity activity until the weights on the periphery of the Gaussian become insufficient to activate the cortical columns, corresponding to this limit of 7°.

We observe that the stationary peak delay (SPD) is also affected by the increase in cortical extent, Fig. 6 D. For cortical extent smaller than 2° the cortical peak is delayed with respect to the RGC peak. In particular, as the connections length tends to 0, the SPD difference (black trace) in Fig. 6 tends to 0. This justifies our comment in section 3.2.1 where we claimed that the difference between RGC and VSDI SPD is primary due to the cortical horizontal connectivity. Beyond 2° the cortical SPD decreases (while the RGC SPD obviously stays constant), Fig. 6 D, so that the VSDI signal peak is more and more in advance to the RGC peak with a saturation at about 5°. The time scale of this peak delay (maximum -14ms) is quite lower than the ML though. It is important to note that this effect is weaker than anticipation by latency and does not evolve along the trajectory of the bar, as shown in Figure 6E. This means that the effect is the same whatever the position of the cortical column on the trajectory of the bar. This shift could be independent of latency anticipation.

We have two hypotheses concerning the potential mechanisms leading to this peak shift, both resulting from an adapting-like anticipation mechanism, reducing the maximum amplitude of the response and shifting its peak. The first hypothesis is based on the intrinsic properties of the cortical columns in the model. This is a gain control mechanism. Due to the non linearity of the conductances a change in the baseline activity of the column can induce a faster drop off of the signal, triggering an advanced peak shift. Note that the baseline activity depends on the cortical extents. The second hypothesis, which can cumulate with the previous one, is a based on a network effect, and more precisely on inhibition. When the bar arrives, the inhibitory effects of upstream activated cortical columns lead to an earlier peak drop off. Both hypotheses actually depend on the cortical horizontal connectivity and could explain the peak shift decay as the inhibitory cortical extent decreases.

To conclude, anticipation is enhanced by increasing the length of excitatory and inhibitory fibres at a constant ratio for values inferior to 7° with an earlier shift in the VSDI peak. Unlike anticipation by latency, the peak shift is independent of the trajectory, though. Bothe effects are highly dependent on changes in the Gaussian profile. They are the strongest with an excitatory length of 4° and an inhibitory length of 0.8°. This is close to the physiological connectivity that we use in the rest of the paper.

3.2.4 The role of conduction velocity

We finally investigated the influence of fibre conduction velocity between cortical columns on anticipation in control conditions (Figure 7), in the range $100 - 333\ ^\circ/s$. The effects are rather easy to summarise and are shown in Fig. 7. There is no effect on the VSDI signal. The maximal latency (ML) slightly decreases while the anticipation range (AR) slightly increases. More interesting is the behaviour of short range anticipation range (SRAS) which is increasing almost linearly. In the paper (Benvenuti et al., 2020) Benvenuti et al. proposed a phenomenological, physiologically plausible model of horizontal cortical integration in response to a moving bar (Fig. 3 of their paper). In their model, SRAS increases linearly with a 1:1 ratio to the fibre conduction speed (Fig. 3D of their paper). In our model, which integrates more biological features, we also observe a (quasi)-linear behaviour but the slope, ~ 0.035 is far from 1. In our opinion, this is because they only used horizontal excitation while inhibitory horizontal connectivity also plays an important role. Inhibition acts as in impediment to the activation of cortical columns, explaining the small slope. The increase in fibre conduction speed is also accompanied by a very slight decrease of 0.6ms in the shift between the cortex and the ganglion cells Stationary Peak Delay (SPD) (7 E).

To sum up, the only remarkable effect induced by an increase in the conduction velocity is a linear increase in the SRAS, similar to what has been conjectured by Benvenuti et al in (Benvenuti et al., 2020), but with a quite smaller slope, due to horizontal inhibition.

3.2.5 Conclusion of section 2.4.2

In normal, biological conditions, the input received by the cortex corresponds to a retina having gain control and amacrine connectivity generating non-linear mechanisms of anticipation. Hence, the previous section shows how the cortex shows differences with

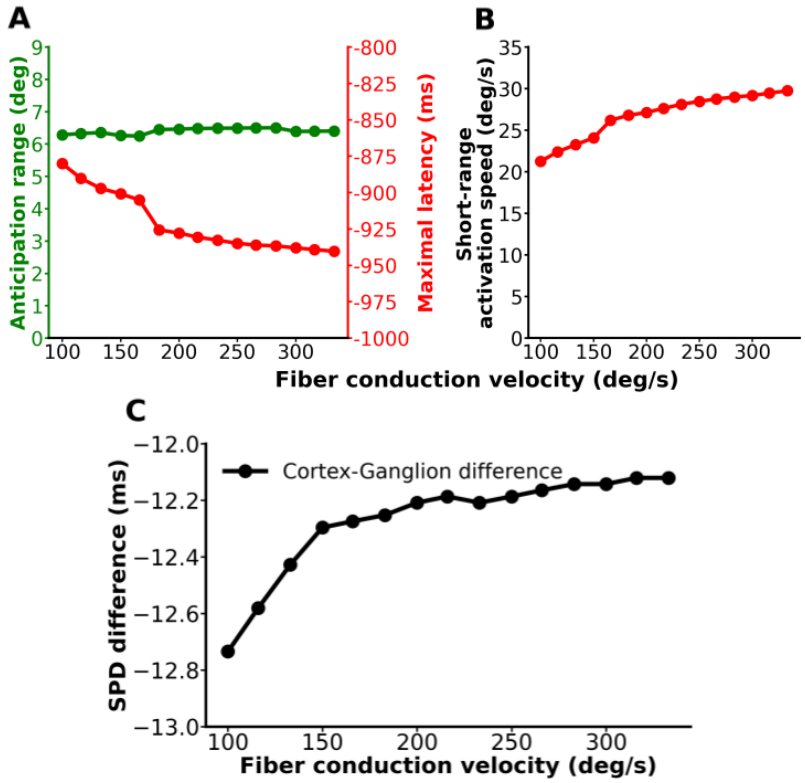

Figure 7: **The effect of fibre conduction velocity on the cortical response A) Temporal and spatial observables:** anticipation range (green) and maximal latency (red) in function of excitatory and inhibitory extent. **B) Speed observable:** short-range activation speed (red) in function of fibre conduction velocity. **C) Difference** between RGC stationary peak delay (SPD) and VSDI signal SPD.

VSDI signal measurements on cortical anticipations in atypical conditions without such mechanisms. We have indeed shown that our cortical model can reproduce the cortical anticipation observed experimentally, although with some quantitative discrepancies. The anticipation in our model clearly results from the propagation of activity in an excitatory/inhibitory balanced horizontal network. We have also shown that, in addition to cortical anticipation latency, the cortical columns exhibit an adapting anticipation mechanism that is independent of latency anticipation and causes an earlier peak offset. This mechanism could be associated with dynamical normalization generated by cortical inhibition and leading to an adapting-like anticipation in the cortex. This is why it is affected by changes in lateral connectivity and ROA. The ability of this system to anticipate is closely linked to a few parameters. Increasing the amplitude (Fig. 3) or the length of connectivity at a constant inh/exc ratio (Fig. 6) improves the ability to anticipate by shift and latency. On the contrary, increasing the speed of the bar (Fig.5) first increases anticipation then decreases it, beyond a "preferred speed". Finally, increasing the conduction velocity of cortical fibres (Fig.7) allows anticipation to propagate more rapidly in the cortex, as expected. This propagation remains limited by inhibition though.

3.3 Cortical anticipation is influenced by the retina

We now investigate how the cortical response behaves if the retinal drive is itself generating peak anticipation during the integration of the motion. Our retina model first implements gain control following (Berry et al., 1999; Chen et al., 2013). This has the effect of advancing the peak response of ganglion cells activity. The effect increases with the size and the contrast of the bar while it decreases with its velocity (Souihel, 2019; Souihel et al., 2019). Anticipation in the retina can also be studied from the point of view of the population, taking into account the interactions between the different cells. We have therefore implemented amacrine cell connectivity and designed a retinal circuit capable, under certain conditions, of improving retinal anticipation. Figures 8 and 9 illustrate these experiments with retinal anticipation in the context of respectively bipolar gain control or feedforward amacrine inhibition. Further investigations can be found in the supplementary figures. Theoretical and numerical results concerning the role of lateral connectivity in retinal anticipation of motion trajectories can be found in (Souihel, 2019; Souihel et al., 2019). Note that, in general, these mechanisms generating peak anticipation also modify the time course of the RGCs response to the moving bar (see Fig. 8 I, 9 I). The resulting effect on the cortical response is thus not only a shift in the VSDI signal peak, but a change in its global shape as well (see e.g. Fig. 8 J, 9 J or Fig. 10). This entails potential changes in the VSDI signal amplitude and in the latency.

Gain control and amacrine cells connectivity cause, on one hand, a decrease in the Retinal Output Amplitude (ROA), and on the other hand, a shift of the response peak earlier while changing its shape. Now, as we saw above, decreasing the ROA impacts the cortical response. Thus, to isolate the effect of e.g. gain control on anticipation we need to compare the situation with gain control to the situation without it, while the ROA are identical. The set of control simulations where no effect (gain control or amacrine cells network) is present, but where the ROA is rescaled to match the case where the effect is present is called "equivalent retinal output amplitude" (EROA).

3.3.1 Gain control in the retina enhances cortical anticipation

In the model, gain control can be present in BCs or RGCs. We ran simulations with increasing the strength of these two gain controls and compare these results to EROA conditions. We only show here the figure for bipolar gain control, which is the most representative; the figure for ganglion gain control can be found in the supplementary figures.

Variation of BCs gain control. We first simulated the response to the moving bar with a bipolar gain control (parameters h_B in eq. (2)) varying between 0 and 9.2mV/s (Fig. 8). As shown in (Souihel, 2019; Souihel and Cessac, 2019) gain control triggers the appearance of adapting anticipation in the cortex, as defined in section 2.4.1. We investigate here what is the induced effect on cortical anticipation, by latency and by peak shift.

The first difference between the case with bipolar gain control ($h_B = 9.2$ mV/s) (Fig. 8 B) and the control case (Fig. 8 A) is the presence of a second peak in the VSDI signal response. This peak shows an abrupt reduction in the response of the retinal cells to the

point where their gain control is reduced while they are still receiving inputs from the OPL. In addition the latency are slightly smaller than in control conditions. The increase in h_B is also associated to a moderate decrease in the amplitude of the VSDI signal (Fig 8 E, C, D). This was expected since the strength of gain control in BCs reduces the amplitude of the retinal response. We observe however a slight deviation of the VSDI signal compared to the EROA condition 8 E : the amplitude of the VSDI signal remains larger. This is because, when h_B increases, gain control changes the shape of the BC response (data not shown) and thereby the RGC profile integrated by the cortex (Fig. 8 I, J). Figures 8 C and D also show a decrease in maximal latency (ML) and anticipation range (AR) respectively with increasing bipolar gain control: AR and ML are slightly favoured for small values of h_B (≤ 4.3 mV/s) before being attenuated. There is no difference with EROA for AR. In the case of the ML, the decreasing regime is less marked than for EROA condition. For a gain control of 9.2 mV/s, ML increases by 7.6% more than EROA (Fig. 8 F). This indicates an additional positive effect of gain control on ML, which partially compensates for the anticipation decay due to retinal amplitude output reduction. The short range activation speed (SRAS) increases much more than in the EROA condition (Fig. 8 G). In comparison, it is 26.7% faster than the speed observed in EROA for BCs gain control at 9.2mV/s. This effect is strong enough to compensate and even reverse the slight decrease due to the reduction in the ROA.

Concerning the anticipation by peak shift, the presence of BCs gain control results in a -77ms (-51.5%) increase in RGC stationary peak delay (SPD) between $h_B = 0$ and $h_B = 9.2$mV/s (Fig. 8 I). This is accompanied by an earlier shift in the cortex peak of -72.5ms (-53%) (Fig. 8 J), corresponding to a strong anticipation by peak shift. Note the difference in the response profiles for $h_B = 0$ (red traces) and $h_B = 9.2$mV/s (green dashed traces) in the RGC response and in the VSDI response. Fig. 8 H illustrates the evolution of the RGC SPD (orange trace), the cortical SPD (blue trace), and the difference of these two quantities (black trace). This difference shows a maximum at $h_B = 4.9$mV/s.

To sum up, this study demonstrates the direct influence of retinal peak anticipation on the cortex when increasing the BCs gain control. BCs gain control induces an earlier shift of the RGC peak, a reduced ROA, and a change in the RGC response profile, with a corollary anticipation in the VSDI signal, by adaptation and by latency. However, the impact on anticipation depends on the level of BCs gain control. For small h_B (≤ 4.9) the main effect is an increase in the peak shift with no significant effect on the cortical anticipation by latency. For larger h_B anticipatory waves propagate faster (SRAS) than in EROA condition but with a reduced latency (ML) and a reduced range at which cortical columns begin to anticipate (AR). Large h_B lead to a reduction in the peak shift of the cortex, detrimental effects on ML and AR, while SRAS saturates. This suggests therefore that there is an optimal value for BCs gain control.

To conclude this section we would like to comment on an effect not discussed so far. The small "rebound" in the VSDI signal observed in Fig. 8 B, after the first peak. This effect, caused by BC gain control in our model, could be tracked in experimental VSDI signal. If such a rebound was found, it would not necessarily mean that it comes from BCs gain control, but at least that a mechanism of the same nature holds in the pathway

from photoreceptors to V1. This could actually also help calibrating better chimera like models.

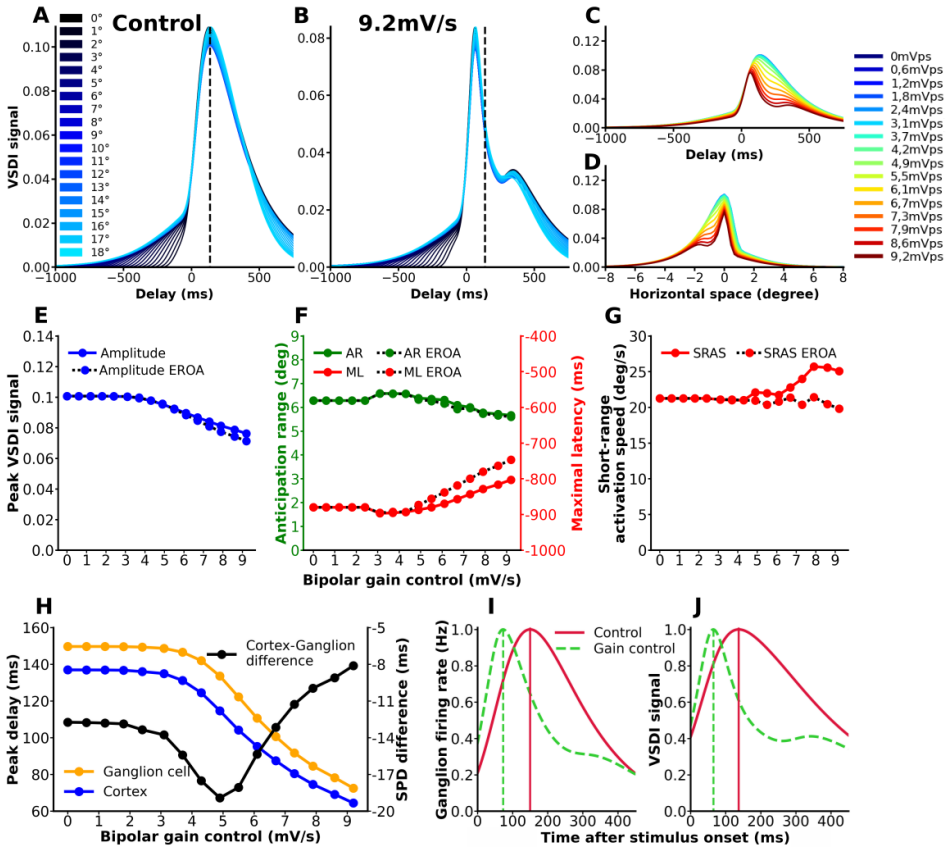

Figure 8: **The effect of bipolar gain control strength, h_B, on the cortical response. VSDI signal response to bipolar gain control at A)** 0 **mV/s (control) and B)** 9.2 **mV/s.** The color bar on the left of Fig. A represents the color gradient introduced in section 2.4. **C) Temporal VSDI signal** in response to increasing bipolar gain control for the cortical column located at the center of the layer ($x = 9°, y = 1.35°$). **D) Spatial VSDI signal** in response to increasing bipolar gain control, for the time where the central cortical column reaches its maximum. The x coordinates has been shifted so that the central cortical column is actually located at $x = 0$. **E) VSDI signal amplitude** of the central cortical column versus h_B. **F) Temporal and spatial observables:** anticipation range (green) and maximal latency (red) versus h_B. **G) Speed observable :** short-range activation speed (red) in function of bipolar gain control weight. In A, B, C, we also drawn the Equivalent Retinal Output Amplitude (EROA) curve. This is the dotted black curve with the same coloured symbols. **H) Stationary peak delay (SPD)** for RGCs (orange), VSDI signal (blue) for the central cell (scales on the left) and difference between RGC SPD and VSDI signal SPD (black, scales on the right). **I) Shape of the central RGC response profile to the moving bar**, without gain control (red) and with BC gain control 9.165 mV/s (dashed green). Note that the two traces have been rescaled to have the same maximum. This is to emphasize the change in the shape of the response induced by BCs gain control. **J) VSDI signal**, same conditions. In I,J, the dotted lines correspond to the peaks in the RGC firing rate or VSDI signal without BCs gain control (red) and with it (green).

Variation of RGCs gain control. We also varied the RGCs gain control, h_G, from 0 mV^{-1} Hz to 0.54 mV^{-1} Hz (Supplementary figure S1). Note that the range of values is very different from BCs gain control because the non linearity in the gain function is quite different (see eq. (5) versus (8)). The result shows strong similarities with BCs gain control though.

Increasing ganglion gain control has similar effects to those observed for bipolar gain control. RGCs gain control induces as well an earlier shift of the RGC peak, a reduced ROA, and a change in the RGC response profile, with a corollary anticipation in the VSDI signal, by adaptation and by latency. We also have less cortical anticipation through latency : the maximal latency (ML) and the anticipation range (AR) decrease while short-range activation speed (SRAS) increases before decreasing. ML and SRAS decrease less than in EROA, which means that the ganglion gain control generates a compensation of the ROA decay. There are three main differences with BCs gain control, though. We first observe an absence of increase in AR and ML for low values of ganglion gain control (below 3 mV/s). In addition, the difference between the time to cortical peak and retinal peak increases slightly before becoming constant while for BCs it increases below 5mV/s before to decrease. Finally, in contrast to BCs gain control (Fig. 8 B) there is no rebound in the VSDI signal after the first peak (see the comment at the end of the previous section).

To summarise, we mentioned in section 3.1 that the time of the peak in CTL conditions is located at 139 ms, while the peak observed in experiments (Fig. 4D of [8]) is very close to 0. We proposed that the discrepancy ought to be due to the absence of retinal anticipation mechanisms in CTL conditions. Here we see that combining the maximal peak shift due to BCs gain control (-72 ms) and the peak shift due to RGCs gain control (-52 ms) we arrive at a cumulative peak shift of -134 ms, so that, compared to CTL, the time to peak with gain control is close to 0, as observed in experiments. Although, the combined effect of bipolar and ganglion gain control can be non linearly entangled, this simple summation provides a fairly good agreement with experimental observations.

3.3.2 The anticipatory role of amacrine cells

Here, we study the effect of lateral ACs inhibition. ACs, like horizontal cells, provide a lateral connectivity entangling the "vertical" information channels from photoreceptors to retinal ganglion cells via bipolar cells. Although some ACs can have excitatory connections (e.g. cholinergic) we mainly focus here on inhibitory effects of ACs. Amacrine cells constitute networks which modulate and can potentially propagate the response of BCs to a moving signal to other BCs and RGCs. This depends though crucially on the scaling between synaptic weights (w_B^A, w_A^B, w_G^A, w_G^B in the model) and characteristic integration times (τ_A, τ_B, τ_G in the model). A detailed study in (Kartsaki et al., 2024) shows that tuning these parameters dramatically change the shape of the response to simple flashes (e.g. from monophasic response to biphasic), while (Souihel, 2019; Souihel and Cessac, 2019) emphasize the effect of stimulus induced wave propagation leading to an advancement of the peak time in RGCs. The network response to the moving bar also depends on the connectivity structure. By construction, our model has a feedback connectivity where BCs act on ACs which modulate BCs back.

But, one can also study a feedforward case where BCs input ACs without reciprocal connection, simply by setting $w_B^A = 0$. S. Ebert (Ebert, 2023) shown in her thesis that the response is rather different. Especially, with feedback connections ACs can induce effects similar to BCs gain control while modifying the spatio-temporal BC response (e.g. leading to biphasic or polyphasic responses even if the OPL input is monophasic (Kartsaki et al., 2024)). The aforementioned studies were focusing on anticipation at the retinal level. Here, we analyse the impact on cortical anticipation, along the same lines as the previous sections, feedforward versus feedback inhibition, although we didn't make a systematic study of these effects when varying the control parameters. As a matter of fact, our conclusions here are limited to one set of parameters values providing examples of what could happen in the early visual system with ACs lateral connectivity takes place.

Amacrine feedforward inhibition. We study first the direct influence of feedforward inhibition pathway with connectivity from BCs to ACs and from ACs to RGCs, respectively characterized by the parameters w_A^B and w_G^A in eq. (2). We varied $w_G^A \leq 0$, controlling the intensity of ACs synapses to RGCs, and $w_A^B \geq 0$, controlling the intensity of BCs synapses to ACs, from 0 to 1 Hz, restricting to the case $w_G^A = -w_A^B$. The other parameters are tuned to the control value (appendix 5.1). The results are shown in Fig. 9.

In the presence of amacrine connectivity strength $\left| w_G^A \right| = w_A^B = 12$ Hz (9 B), the VSDI response as a function of distance from the origin of the bar shows a very reduced latency and a very different shape of the curve, narrower and accompanied by a second small peak separated from the first compared with the control (9 A). As expected, the inhibitory effect of ACs induces a decrease in the VSDI signal amplitude (Fig. 9 E, C, D) but also in maximal latency (ML) (Fig. 9 C) and anticipation range (AR) (Fig. 9 D). Note that the VSDI signal curve in this condition is essentially indistinguishable from the equivalent retinal output amplitude (EROA) curve (Fig. 9 E), confirming that the effect of ACs on VSDI, in feedforward conditions, is only a drop in the ROA. Fig. 9 F shows also a detrimental effect of feedforward inhibition on the anticipation range (AR), as well as on the maximal latency (ML), more important than in EROA. In contrast, the comparison of the short range activation speed (SRAS) obtained with feedforward inhibition and its corresponding EROA reveals a compensatory effect of the retinal amplitude. While SRAS at EROA decreases brutally, the SRAS associated to feedforward amacrine connectivity remains relatively constant suggesting a mechanism which counterbalances the decrease in speed due to the decrease in amplitude.

The most prominent effect occurs on the Stationary Peak Delay (SPD) of RGCs (orange trace) and VSDI signal (blue trace) which decrease sharply (Fig. 9 H), though with a difference becoming smaller and smaller in absolute value (black trace). These behaviors are accompanied by a strongly advancement of -104 ms (-69.4 %) for the RGC SPD (Fig. 9 I) and of -88 ms (-64.2 %) for the VSDI signal peak (Fig. 9 J), when $\left| w_G^A \right| = w_A^B = 1$ Hz. Note that we observe also a small rebound of the VSDI signal (Fig. 9 J), arising when $\left| w_G^A \right|$ becomes large. When the bar arrives in their receptive field, BCs activity increases, increasing the ACs activity and their inhibitory effect on the RGC. When $\left| w_G^A \right|$ is large enough it takes over the excitation from BCs and the RGC firing rate drops to 0. Because the ACs have here a shorter time scale than BCs

(corresponding e.g. to the peak of the green dashed curve in Fig. 9 I), their effect lasts shorter and the rebound corresponds to the residual activity coming from BCs.

Thus, the presence of feedforward inhibitory connectivity implements adapting anticipation within the cortex, with an increasing effect as the intensity the synaptic weights increases, while severely penalising cortical latency anticipation (AR, ML) and intrinsic cortical shift. Surprisingly, though, the speed of the anticipatory wave remains stable thanks to compensation for the decay of the ROA. All these effects are not simply due to a decrease in the ROA but entail additional effects presumably due, as for gain control, to the difference in the BCs response profile.

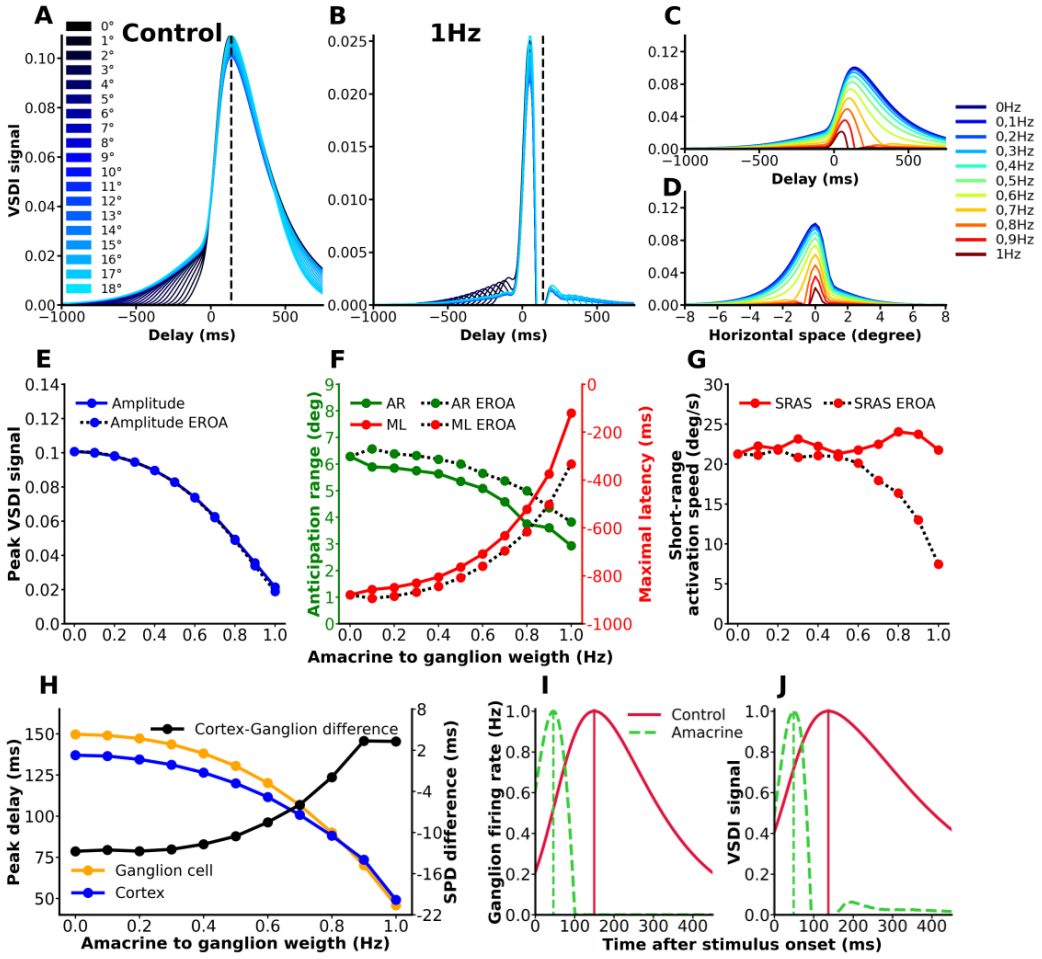

Figure 9: **The effect of feedforward inhibition, controlled by $w_G^A = -w_A^B$, on the cortical response. VSDI signal response to feedforward AC connection strength at A)** 0 **Hz (control) and B)** 1 **Hz.** The color bar on the left of Fig. A represents the color gradient introduced in section 2.4. **C) Temporal VSDI signal** in response to increasing feedforward AC connection strength for the cortical column located at the center of the layer ($x = 9°, y = 1.35°$). **D) Spatial VSDI signal** in response to increasing feedforward AC connection strength, for the time where the central cortical column reaches its maximum. The x coordinates has been shifted so that the central cortical column is actually located at $x = 0$. **E) VSDI signal amplitude** of the central cortical column. **F) Temporal and spatial observables:** anticipation range (green) and maximal latency (red). **G) Speed observable :** short-range activation speed (red). **H) Stationary peak delay (SPD)** for RGCs (orange), cortical columns (blue) and difference between the cortical and RGC SPD (black). **J) Shape of the central RGC response profile to the moving bar**, without feedforward AC connection ($w_G^A = -w_A^B = 0$) (red) and when these parameters take the maximum value ($\left| w_G^A \right| = w_A^B = 1$ Hz, dashed green). Note that the two traces have been normalized to have the same maximum at 1 Hz. **I) VSDI signal**, same conditions. The dotted lines in E, F correspond to the peaks in the RGC firing rate or VSDI signal without feedforward AC connection (red) and with it (green).

Amacrine feedback inhibition. The feedback loops BCs $\rightarrow$ ACs $\rightarrow$ BCs induce a complex interaction between the moving object, the local cells response and the influence of these cells on distant cells, propagated via the lateral amacrine inhibition. The picture is that of a moving bar propagating in a non homogeneous landscape of activities modulated by the amacrine network. This entails specific effects, not present with a feedforward connectivity, such as the existence of resonant frequencies (Cessac, 2022), a change in the shape of BCs response (Kartsaki et al., 2024) or the existence of a preferred speed range at which anticipation by adaptation is maximal (Ebert, 2023). In addition, the feedback loop propagates inhibition producing a decay of the response before and after the peak. This actually differs from gain control, which only acts on the post-peak portion of the curve. It has been shown that these feedback effects can be characterized by (1) the characteristic integration times of BCs (τ_B) and ACs (τ_A); (2) the average intensity of the connection from BCs to ACs, $w_A^B \geq 0$, and ACs to BCs, $w_B^A \leq 0$. Actually, the relevant parameter is $\mu = -w_B^A w_A^B \tau^2$, where $\frac{1}{\tau} = \left| \frac{1}{\tau_B} - \frac{1}{\tau_A} \right|$. In particular, this shows that the synaptic weight effects of the feedback loop is characterized by the product $w_B^A w_A^B$. Here, we keep τ_A, τ_B constant and vary w_A^B with $w_B^A = -w_A^B$ from 0 to 12 Hz (Supplementary figure S2). The values of the other parameters are given in the Appendix 5.1. The goal here is to show an example of the potential effects of this type of connectivity while a more exhaustive study would require to vary other parameters as well, a task well beyond the scope of this paper.

These results demonstrate the capacity of amacrine retro-control to provoke an increasing adapting anticipation influencing the cortex. However, this is accompanied by a negative effect on cortical anticipation by latency mainly manifested by a drop in the AR, ML and of the cortical peak shift. In contrast, the anticipatory wave speed is increasing a little bit. In general, these effects were much less pronounced than for feedforward inhibition, and no second peak was observed in the time course of the VSDI.

3.3.3 The anticipatory impact of the retina

We now synthesize our observations about the cortical correlates of the retinal anticipatory mechanisms. Although our investigations are absolutely non exhaustive - a more detailed study would require to vary a larger set of parameters in the model - it reveals several effects which are generic, i.e. present on a wide range of parameters value, although the *quantitative* observations may depend on these parameters. These generic effects are an advancement of the RGC output peak, a decrease in the ROA, and a global change in the shape of the firing rate response. This has an impact on the VSDI response, illustrated in the figures above and summarized in Fig. 10.

In Fig. 10 A, we compare the respective effect of BC and RGC gain control on the VSDI signal at the central cortical column. It is however rather difficult to compare quantitatively the effects of these two gain control mechanisms, as a systematic study would require to modify jointly h_B and h_G in a two dimensional map, a task beyond the scope of this paper. Here, we only compare the model response in a case where BCs and RGCs gain control are tuned so that the retinal output amplitude is the same (ROA $= 17.1$ Hz), that is $h_B = 8.554$mV/s (magenta trace) and $h_G = 0.359$mV/s (yellow trace). The first observation is a peak shift stronger for BC gain control than for RGC gain control.

There is a SPD difference of 25ms (36.7%). There is also a strong difference in the VSDI signal profile, after the pic. It is unfortunately not possible to compare this with the experimental results in (Benvenuti et al., 2020) as the VSDI profiles they show (e.g. Fig. 4 D) is cut before the place where such a possible rebound may occur. Performing new experiments on a larger time scale would be a way to confirm the role of BC gain control on the VSDI signal profile. Finally, one sees a small variation in the early part of the curves, where latency is computed. The smallness of this variation is due to the scaling of the figure though, where we wanted to show the whole VSDI profile. Now, remember that cortical latency observables are computed at the very beginning of the activity rising, when this activity exceeds a threshold of 0.001 (section 2.4). Thus, there is a small but significant effect on latency observables. There is a difference of -1.2% for AR , -2% for ML and -7% for SRAS, in favour of BC gain control. This suggests that, overall, BC gain control is more effective. However, cortical anticipation by latency is slightly affected compared to adapting anticipation which is much higher. To sum up BCs control appears more effective at generating peak anticipation while limiting the impact on cortical latency anticipation and increasing the speed of anticipation.

In the same vein, Fig. 10 B compares the role of feedforward versus feedback amacrine cells connectivity, adjusting the control parameters $w_G^A = 0.7$ Hz (feedforward) and $w_B^A = 11$ Hz (feedback) so that the ROA are equal to 13.1 Hz. We observe here a small pic shift and a small change in latency while the main visible effect is a change in the shape of the VSDI profile after the peak. For these value of parameters, the AR actually decreases by -6.9% when comparing the feedback case to the feedforward case, the ML decreases of -6% and SRAS by -7.3 %. Finally, SPD arrives 9ms later in the feedforward case than in the feedback case $(10.6$ %). Although the difference is rather tiny, a general conclusion would require a more systematic study. In particular, a comparison of Fig. 9 and Fig. S2 suggests a better efficiency of amacrine connectivity in the form of a feedback loop. Adaptating anticipation is more pronounced while limiting the impact on cortical latency anticipation despite the smaller intrinsic cortical delay. On the basis of preliminary results obtained in (Ebert, 2023) we actually believe that the main difference between the two effects would hold when varying the bar speed. We expect the presence of a preferred bar speed - where anticipation is maximum - in the feedback case and not in the feedforward case (where anticipation would grow monotonously until saturation).

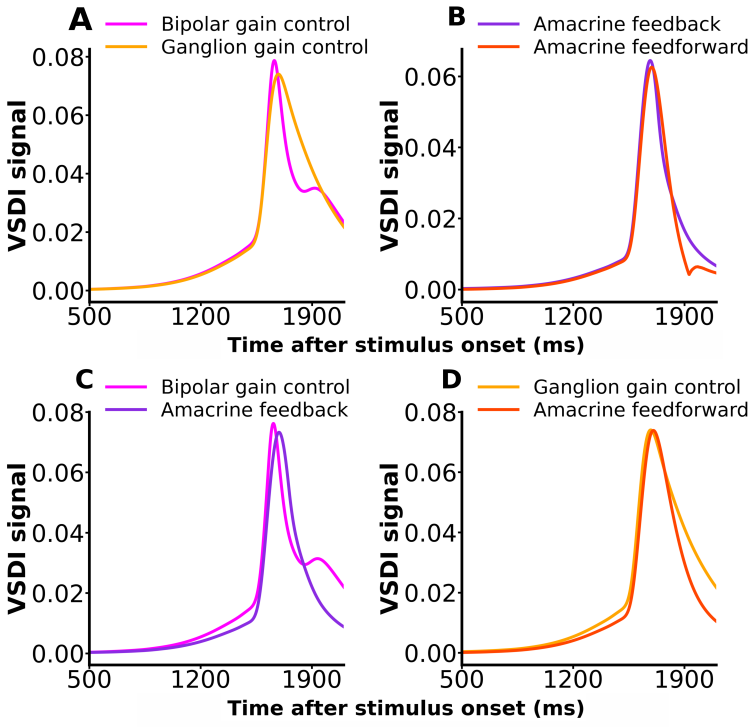

Figure 10: **Gain control and amacrine connectivity comparison. A) Time course** of the VSDI response with a BC gain control $h_B = 8.554$ mV/s (magenta) and a RGC gain control $h_G = 0.359$ mV/s (yellow). **B) Time course** of the VSDI response with amacrine feedback connectivity weight $w_B^A = 11$ Hz (purple) and feedforward weight $w_G^A = 0.7$ Hz (orange). **C) Time course** of the VSDI response with bipolar gain control $h_B = 8.554$mV/s (majenta) and amacrine feedback connectivity weight $\left| w_B^A \right| = 11$ Hz (purple). **D) Time course** of the VSDI response with ganglion gain control $h_G = 0.359$mV/s (yellow) and amacrine feedforward connectivity weight $\left| w_G^A \right| = 0.7$ Hz (orange).

Fig. 10 C compares the BCs gain control to the ACs feedback still tuning the respective parameters so that the ROA are equal: BCs gain control $h_B = 9.2$mV/s and the feedback loop $\left| w_B^A \right| = 9$ Hz (same ROA of 16.1 Hz). Considering the difference between AR in the BC gain control case and in the feedback connectivity case, we obtain a decrease of -8.4 %. This is -10.3% for ML, and -4.9% for SRAS, while SPD arises 36ms later 55.2% and the intrinsic cortical SPD 3.4ms (43%) after. In this example BCs gain control provides more anticipation with a visible effect on latency. Fig. 10 D compares as well the RGCs gain control to the ACs feedforward connectivity with $h_G = 0.36$mV/s, $w_A^B = 0.6$ Hz (ROA equal to 17 Hz). We observe a decrease of -10.1 % for AR, -11.3% for ML and -8.4 % for SRAS. The RGC SPD arises 17 ms (18.1 %) later and the cortical SPD 6ms (41.4 %) later. For this set of parameters RGC gain control performs better than feedforward amacrine connectivity.

To sum up, this study provides an effective way of studying the potential impact of retinal anticipation mechanisms on cortical anticipation with a main drawback: parameters

tuning. Although, the retinal model has been designed to have a minimal set of parameters (compared to a real retina) there are still quite a lot and a systematic study of the effects requires actually an (ongoing) systematic mathematical analysis (Cessac-Ebert, in preparation). Note that numerical simulations do not allow to effectively sample the parameter space of the retina model, while a mathematical analysis shows that some parameters (such as $\tau_A, \tau_B, w_B^A, w_A^B$) are actually dependent. Another alternative for parameter tuning would be experimental tests. Experiments on the retina somewhat allow one to tune these parameters so as to match empirical responses. But, as said in the introduction, they mainly hold for mice (or salamanders). A more efficient way and a bigger challenge would be to tune these parameters from the observation of the V1 VSDI signal, on the basis of this model and of the afferent simulation platform. We actually made in this paper a conjecture about the shape of the VSDI signal (presence of a rebound after the main peak) when, e.g., BCs gain control or ACs feedback connectivity is present. Their influence results in a mesurable effect on the VSDI signal profile. It would be interesting to test these conjectures in new experiments.

In this spirit, we would like to come to the remark made in section 3.1. In CTL condition (without retinal anticipation) the anticipation by latency is a bit too high (-880 ms for ML) compared to $\sim -400, -600$ ms in Fig. 4D of (Benvenuti et al., 2020) (where this range of variations comes from feedback effects). We have actually shown that adding a realistic retinal output has the effect of reducing the anticipation by latency in our model, in a range compatible with experimental observations (see e.g. Fig. 9 B). The same holds for the anticipation range which was 6.3° in CTL conditions compared to 2° in experiments. Again, the retina has the effect of reducing this discrepancy. To get better insights on the parameters values experiments could for example focus more on the VSDI shape (e.g. after the peak) and also investigate the effects of the bar speed on peak anticipation and latency anticipation.

4 Discussion

In this paper, we presented a chimera model connecting a multi-layered model of the retina to a mean-field model of V1. This allowed us to investigate how the retino-cortical pathway (where the thalamus-LGN was considered as a simple relay) can anticipate the trajectory of a moving object. Cortical anticipation is first manifested by a spatial gradient in the latency of the VSDI signal. This corresponds to a cooperative and cumulative non linear mechanisms where the cortical columns sensing the bar send information feedforward to pre-activate distant columns, generating what we called "anticipation by latency". We have also exhibited an advancement of the cortical peak, compared to the peak of the retinal ganglion cells that input cortical columns. In contrast, the time scale of this shift is smaller than that of the latency and is not affected by the position along the motion trajectory. It corresponds to anticipation by adaptation supported by cortical inhibition. We therefore have a parallel effect of inhibitors in the retina and cortex, generating adapting anticipation. Considering a simple retinal input (essentially an α-drive as in (Benvenuti et al., 2020)), with no anticipatory mechanism, our results highlight what happens in the cortex in the absence of retinal anticipation. We have shown that the amplitude of the retinal output or the stimulus contrast, as well

as the length of horizontal cortical connections, have a strong impact on the cortical anticipation by latency. We also studied the influence of the moving object speed where we gave evidence of a preferred speed where the anticipation by latency is maximal. In control conditions, the effects on the peak shift are quite smaller than the effects on latency.

We have then study cases where the retina actively cooperates with the cortex V1 to anticipate the trajectory, as in biology. Essentially, the retina mechanisms that we have investigated (gain control and amacrine connectivity) induce a strong advancement of the retinal ganglion cell response peak, with a potential deformation of the whole retinal signal when the mechanism modulates the activity of bipolar cells. In the presence of retinal anticipation mechanisms, the retinal peak shifts feedforward and the response profile deformation generates entangled effects in the cortex. Not only the VSDI signal peak is advanced and is located at a position that matches experimental observations. Its whole shape is modified as well, leading to measurable and various effects on anticipation by latency. Depending on the control parameter value this can improve or decrease the anticipation range and the maximal latency, while the speed at which the anticipatory wave is propagating (short range anticipation speed) is a bit increased or stays constant. Although the mechanism of peak anticipation is known since long in the retina, (Berry et al., 1999), our work is the first, to our knowledge, showing how the effect of the peak offset in the retina is transmitted to the cortical response.

This study is far from being exhaustive though. First, a more extended study would require to vary the phenomenological parameters controlling such or such mechanism (e.g. the parameter h_B for bipolar gain control or the parameter $\mu = -w_B^A\, w_A^B\, \tau^2$, where $\frac{1}{\tau} = \left| \frac{1}{\tau_B} - \frac{1}{\tau_A} \right|$ for feedback amacrine connectivity) along with the bar speed. Indeed, preliminary studies on the retina (Ebert, 2023) have shown that acting on bipolar cells anticipation induces the existence of a speed range (preferred speeds) where anticipation by adaptation is maximum. What would be the effect on cortical anticipations ? Is there also a range of preferred speed, and do they correspond to physiological or psychophysical observations ? Such an extensive study would also require to explore more deeply the set of parameters involved in the cortical model in future studies. As we briefly explained in the paper, this model exhibit pathological instabilities, when the retinal output amplitude increases too much. These instabilities can be controlled either by considering a so-called second order mean-field model (ElBoustani and Destexhe, 2009; Aquilué-Llorens et al., 2024) or an adaptive mean-field model (Di Volo et al., 2019). In particular the second order model introduces corrections in the Jacobian of the cortical model at the steady state. These corrections increase the value at which the retinal output amplitude triggers a Hopf bifurcation generating pathological oscillations. An important improvement would also be to integrate a realistic thalamus in the model, that is, adding thalamic mean-field equations so as to design a mean-field model of the retino-thalamo (LGN)-cortical pathway (Overwiening et al., 2024). Most excitatory synapses in the thalamus are of cortical origin (Sherman and Guillery, 2011), and thus the cortico-thalamic feedback, a feature not present in our model, is essential. In addition, thalamic neurons have very complex intrinsic firing properties, and can generate bursts of action potentials, so they are also not a simple relay. A future step would be to integrate

these properties of the LGN by using a previously proposed model of the thalamus using AdEx neurons, which are formulated in mean-field (Destexhe, 2009) and integrated in our simulator. This would lead us to simulate the role of the thalamo-cortical loop not only for anticipation.

The retina model could be extended through an implementation of other cell types and more diverse types of connectivity, more realistic than nearest neighbours (Souihel and Cessac, 2019). This would allow to implement more realistic retinal circuits, and assess their role in motion processing and anticipation. It would be in particular interesting to study the combined effect of gap junctions and amacrine cells connectivity, and assess whether their collaboration can improve cortical anticipation as suggested by (Souihel and Cessac, 2019).

The current implementation of our retino-cortical model takes as an input a set of images, processes them through the layered retina model in order to produce firing rates, which are then fed to the cortical model. It is hence possible to study the response of the model to different kind of 2D stimuli, and assess whether the mechanisms we have implemented for motion anticipation can also account for other visual effects. In particular is there anticipation on curved trajectories ? See (Souihel, 2019) for preliminary results.

Finally, these results actually ask the question which features of the cortical activity profile (VSDI) are most used by the real brain to anticipate a trajectory: latency ? peak shift ? a combination of both ? or signal features that we didn't explore ? In (Menz et al., 2020) Menz et al observed a rise in the RGCs activity before a moving object enters their receptive field, advancing the *entire* retinal response (not only the peak), with a strong anticipatory effect. This response shift is called *predicting anticipation*. The authors hypothetized that this mechanism is due to the lateral inhibition by biphasic ACs, a feature that we didn't explore here. These cells have a first negative part of their activity that disinhibits BCs, followed by a second positive part that inhibits them. These two parts are relatively symmetrical and very close in amplitude. Sending such a signal to V1 would result in an increase in the early part of the peak and a similar reduction in the late part. This would result in a shift of the VSDI response arising earlier with a constant amplitude at the peak, since we have equivalent stimulation and inhibition. Preliminary results in this direction can be found in (Emonet, 2024). This opens up questions about how the various structure of the retina, with multiple cells sub-types and a large palette of synapses, contributes to motion anticipation and more generally to handling motion. In particular, the retina has the ability to detect surprise (Schwartz et al., 2007; Ebert et al., 2024; Despotović et al., 2024), e.g. an abrupt change in a trajectory. How is the surprise information transmitted to the visual cortex, how is it used, and how could it be measured? These are all interesting questions that could be explored in future studies.

Acknowledgments. This work was supported by the National Research Agency (ANR), in the project "Trajectory", `https://anr.fr/Project-ANR-15-CE37-0011`, funding Selma Souihel's PhD; the ANR too in the project "Shooting Star-15755" `https://anr.fr/Projet-ANR-20-CE37-0018`, funding Jérôme Emonet's PhD, and finally by the interdisciplinary Institute for Modelling in Neuroscience and Cognition (NeuroMod http://univ-cotedazur.fr/en/idex/projet-structurant/neuromod) of the Université

Côte d'Azur. Mattéo Di Volo and Alain Destexhe are supported by the CNRS and the European Union (The Human Brain Project, H2020-945539). We thank O. Marre and S. Chemla for helpful discussions. We are grateful to the Service d'Etudes et Développement at Inria for helping develop the software Macular.

5 Appendix

5.1 Model parameters (default values)

Parameter category	Parameter Name	Symbol	Value	Unit
Simulated area	Number of cells (X)	n_{cells_X}	83	Dimensionless
	Number of cells (Y)	n_{cells_Y}	15	Dimensionless
	Width (X)	L_X	18.45	degrees
	Height (Y)	L_Y	3.15	degrees
	Cell distance	δ	0.225	degrees
	mm retina per degree	rpd	0.3	mm/degrees
	mm cortex per degree	cpd	3	mm/degrees
Stimulus video (moving bar)	Speed	v_B	6	degrees/s
	Width (X)	l_X	0.67	degrees
	Height (Y)	l_Y	0.9	degrees
	Frame rates	δ_t	60	Hz
	Pixel per degree	ppd	300	pixel/degrees
OPL	OPL input amplitude	$\mathcal{C}$	0.025	Dimensionless
	Integration time step (ODE solver)	dt	0.0004	s
	RF size	σ_c	0.2	degrees
	Characteristic time	τ_c	0.1	s
Bipolar cells	Characteristic time	τ_B	0.1	s
	Activity characteristic time	τ_{a_B}	0.1	s
	Threshold	θ_B	0	mV
	Activity rate	h_B	[0,9.2]	mV^{-1}Hz
Amacrine cells	Characteristic time	τ_A	0.05	s
	Weight bipolar to amacrine	$\left\| w_A^B \right\|$	[0,12]	Hz
	Weight amacrine to bipolar	w_B^A	[0,12]	Hz
	Weight amacrine to ganglion	w_G^A	[0,1]	Hz
Ganglion cells	Radius bipolar to ganglion	σ_G	0.3	degrees
	Characteristic time	τ_G	0.1	s
	Activity characteristic time	τ_{a_G}	0.189	s
	Threshold	θ_G	0	mV
	Activity rate	h_G	[0,0.54]	mV^{-1}Hz
	Weight bipolar to ganglion	w_G^B	0.15	Dimensionless

	Pooling extent	σ	0.09	mm
	Non linear rectification	α_G	1110	Hz/mV
	Non linear rectification limit	$N_{G_{max}}$	212	Hz
Cortical shared parameters values	Membrane capacitance for cortical columns	C_m	0.2	nF
	Excitatory reversal potential	V_E	0	mV
	Inhibitory reversal potential	V_I	-80	mV
	Leak reversal potential	V_L	-65	mV
	Cell number in network	N_{tot}	10 000	Dimensionless
	Excitatory quantal conductance	Q_E	1.5	nS
	Leak conductance	g_L	10	nS
	Excitatory characteristic time	τ_E	0.005	s
	Inhibitory characteristic time	τ_I	0.005	s
	External drive	ν_{ext}	2	Hz
	Mean field characteristic time	T	0.005	s
	Fraction of inhibitory cells	g_{ei}	0.2	Dimensionless
	Probability of connectivity	p_{connec}	0.0375	Dimensionless
	Mean voltage initial fit value	μ_V^0	-60	mV
	Mean voltage initial fit variation	$\delta\mu_V^0$	10	mV
	Sigma initial fit value	σ_V^0	4	mV
	Sigma initial fit variation	$\delta\sigma_V^0$	6	mV
	Normalized tau initial fit value	$(\tau_V^N)^0$	0.5	Dimensionless
	Normalized tau initial fit variation	$(\delta\tau_V^N)^0$	1	Dimensionless
Cortical excitators (RS)	Sigma extent	σ_E	1.67	degrees
	Initial activity	ν_{E_0}	1.86	Hz
	Inhibitory quantal conductance	Q_I	3	nS

	P parameters	P_E	[-49.8, 5.06, -25, 1.4, -0.41, 10.5, -36, 7.4, 1.2, -40.7]	mV
Cortical inhibitors (FS)	Sigma extent	σ_I	0.3	degrees
	Initial activity	ν_{I_0}	12.66	Hz
	Inhibitory quantal conductance	Q_I	5	nS
	P parameters	P_I	[-51.4, 4, -8.3, 0.2, -0.5, 1.4, -14.6, 4.5, 2.8, -15.3]	mV
Cortical column connectivity	Retino-cortical amplitude	w_{RC}	2.5	Dimensionless
	Cortical density	ρ_{cort}	4000	mm^{-2}
	Retinal density	ρ_{ret}	400	mm^{-2}
	Fiber conduction velocity	v_C	300	mm/s
	Weight excitatory to excitatory	A_E^E	1	Dimensionless
	Weight inhibitory to inhibitory	A_I^I	1	Dimensionless
	Weight excitatory to inhibitory	A_I^E	1.5	Dimensionless
	Weight inhibitory to excitatory	A_E^I	1	Dimensionless

Table 1: **Model parameters.** For each parameter in our retino-cortical model, we give its name, symbol, unit and the default value used in control condition.

5.2 Mean-field cortical equations

We summarize here the mean-field equations derived in a series of paper ElBoustani and Destexhe (2009); Zerlaut et al. (2018); Chemla et al. (2019); DiVolo et al. (2019); Di Volo et al. (2019); DiVolo and Férézou (2021). The goal is to propose dynamical equations characterising the average dynamics of the two populations of neurons, exciters and inhibitors, at the level of a cortical column. In the core paper, we have given the main equations ruling the dynamics. Here, we give more detail of their constitutive elements. For clarity we rewrite the dynamical equations (9), giving the evolution of the excitatory population rate, $\nu_E(\vec{x}, t)$, for the cortical column located at $\vec{x}$ at time t, (resp. $\nu_I(\vec{x}, t)$ for

the inhibitory population rate), in the form:

$$\begin{cases} T\,\frac{\partial \nu_E(\vec{x},t)}{dt} \;=\; -\nu_E(\vec{x},t) \;+\; F_E\left[\,\mathcal{I}_E^E(\vec{x},t),\,\mathcal{I}_E^I(\vec{x},t)\,\right] \\ \\ T\,\frac{\partial \nu_I(\vec{x},t)}{dt} \;=\; -\nu_I(\vec{x},t) \;+\; F_I\left[\,\mathcal{I}_I^E(\vec{x},t)\;,\,\mathcal{I}_I^I(\vec{x},t)\,\right]. \end{cases} \tag{12}$$

where:

$$\begin{cases} \mathcal{I}_E^E(\vec{x},t) \;=\; \nu^{aff}(\vec{x},t) \;+\; \nu^{drive} \;+\; A_E^E\,\nu_E^{input}(\vec{x},t), \\ \mathcal{I}_E^I(\vec{x},t) \;=\; A_E^I\,\nu_I^{input}(\vec{x},t), \\ \mathcal{I}_I^E(\vec{x},t) \;=\; \nu^{aff}(\vec{x},t) \;+\; \nu^{drive} \;+\; A_I^E\,\nu_E^{input}(\vec{x},t), \\ \mathcal{I}_I^I(\vec{x},t) \;=\; A_I^I\,\nu_I^{input}(\vec{x},t). \end{cases} \tag{13}$$

Here, $\mathcal{I}_X^Y(\vec{x},t)$, $X,Y = \{\,E,I\,\}^2$, is the total contribution (firing rates) of population Y controlling the time evolution of population X.

Note a few important differences with the aforementioned papers. First, the retino-thalamic input ν^{aff} is, in our case, acting on the two populations, excitatory and inhibitory. Also, we have introduced the coefficients A_X^Y weighting the relative contributions of excitatory and inhibitory populations, $\nu_E^{input}, \nu_I^{input}$, from the other columns. These coefficients were implicitly set to 1 in these papers and, in this case, $\mathcal{I}_E^E = \mathcal{I}_I^E$, $\mathcal{I}_E^I = \mathcal{I}_I^I$ so that the superscript Y was useless.

We then define, dropping the $(\vec{x},t)$ dependence for legibility:

$$\mu_{G_X^Y} = \mathcal{I}_X^Y K_X \tau_X Q_X \tag{14}$$

$$\sigma_{G_X^Y} = Q_X\sqrt{\tfrac{\mathcal{I}_X^Y K_E \tau_X}{2}};\; X,Y = \{\,E,I\,\}^2. \tag{15}$$

where, $\mu_{G_X^Y}$ is contribution of population Y to the mean conductance of population X at $(\vec{x},t)$, and σ_{G_X} the corresponding standard deviation, computed from shot-noise theory Papoulis (1965). K_E (resp. K_I) is the number of excitatory synapses (resp. inhibitory), Q_E (resp. Q_I) the unitary excitatory conductance (resp. inhibitory), and τ_E (resp. τ_I) the excitatory decay (resp. inhibitory). See table in appendix 5.1 for the value of these parameters.

The total input conductance of the neuron μ_G and its effective membrane time constant τ_m^{eff} are controlled by the mean conductances as follows:

$$\begin{aligned} \mu_G \;&=\; \sum_{X,Y\in\{\,E,I\,\}^2} \mu_{G_X^Y} \;+\; g_L, \\ \tau_m^{\text{eff}} \;&=\; \frac{c_m}{\mu_G}, \end{aligned} \tag{16}$$

where C_m is the membrance capacitance, assumed to be the same for all neurons and g_L is the leak conductance.

The transfer functions of excitatory (resp. inhibitory) neurons appearing in eq. (12) take the form:

$$F_X(\mathcal{I}_E^X,\,\mathcal{I}_X^I) = \frac{1}{2\tau_{V,X}}\,erfc\Big(\frac{\nu_{thr,X}^{eff} - \mu_{V,X}}{\sqrt{2}\sigma_{V,X}}\Big), \quad X = E,I. \tag{17}$$

The quantity:

$$\mu_{V,X} \;=\; \frac{\sum_{Y=E,I}\mu_{G_X^Y}\,V_Y \;+\; g_L\,V_L}{\mu_G}, \quad X=E,I, \tag{18}$$

is the mean-voltage of the population X in the cortical column, where V_E, V_I, V_L are respectively the reversal potentials for excitatory (E), inhibitory (I) neurons and for the leak. Likewise:

$$\sigma_{V,X} = \sqrt{\sum_{Y=E,I} K_Y\,\mathcal{I}_X^Y\,\frac{(U_Y\,\tau_Y)^2}{2\,(\tau_{\rm m}^{\rm eff}+\tau_Y)}}, \tag{19}$$

$$\tau_{V,X} = \frac{\sum_{Y=E,I}\left(K_Y\,\mathcal{I}_X^Y\,(U_Y\,\tau_Y)^2\right)}{\sum_{Y=E,I}\left(K_Y\,\mathcal{I}_X^Y\,(U_Y\,\tau_Y)^2/(\tau_{\rm m}^{\rm eff}+\tau_Y)\right)}, \tag{20}$$

where we defined $U_Y = \frac{Q_Y}{\mu_G}(V_Y-\mu_{V,Y})$ and where $X=(E,I)$.

Finally, in eq. (12), $\nu_{thr,X}^{eff}$ is a phenomenological threshold expressed as a first order expansion (eq. (21)) of the three sub-threshold statistical quantities : $\mu_{V,X}$, $\sigma_{V,X}$, $\tau_{V,X}$. As it does not exist an analytic formulation of the transfer function for complex models such as the Adaptive Exponential IF or Hodgkin Huxley, Zerlaut et al. Zerlaut et al. (2018) have developed a semi-analytic method based on a phenomenological threshold which carries the single neuron non-linearities mechanisms (spike/reset and adaptation). This leads to the following expression for the phenomenological threshold:

$$\nu_{thr,X}^{eff} \;=\; P_{X,0} + \sum_{u\in\{\mu_{V,X},\sigma_{V,X},\tau_{V,X}^N\}} P_{X,u}\cdot\left(\frac{u-u^0}{\delta u^0}\right) + \sum_{u,v\in\{\mu_{V,X},\sigma_{V,X},\tau_{V,X}^N\}^2} P_{X,uv}\cdot\left(\frac{u-u^0}{\delta u^0}\right)\left(\frac{v-v^0}{\delta v^0}\right), \tag{21}$$

where $\tau_{V,X}^N = \tau_{V,X}G_l/c_m$ is an unit-less parameter. Coefficients $P_{X,u}$ have been fitted on numerical simulations of a given single AdEx neuron model with conductance-based exponential synapses Carlu et al. (2020). Note that they are different for exciters and inhibitors. The value of these coefficients in our model are given in the table of appendix 5.1, as well as the coefficients μ_V^0, $\delta\mu_V^0$, σ_V^0,$\delta\sigma_V^0$, $(\tau_V^N)^0$, $(\delta\tau_V^N)^0$ which are actually assumed to be constant over the populations E, I.

Finally, the VSDI signal is given by:

$$VSDI(\vec{x},t) = 0.8\;\times\;VSDI_E(\vec{x},t)\;+\;0.2\;\times\;VSDI_I(\vec{x},t), \tag{22}$$

where the coefficient 0.8 and 0.2 corresponds to the fraction of exciters and inhibitors in the colum population and the VSDI is the percentage variation in relation to the base line:

$$VSDI_X(\vec{x},t) \;=\; \frac{\mu_{V,X}(\vec{x},t)-\mu_{V,X_0(\vec{x})}}{\mu_{V,X_0(\vec{x})}}, \quad X=E,I,$$

where $\mu_{V,X_0}(\vec{x})$ is the average membrane potential at rest (i.e. when $\nu^{aff}=0$, for the population X). Note that it depends on $\vec{x}$, due to boundary conditions.

5.3 Connectivity type

Our retino-cortical model contains three different connectivity types :

- **One to one** connectivity link cells or cortical column populations with the same spatial position. This is the case for the connectivity from bipolar to amacrine or from ganglion cell to excitatory/inhibitory cortical column population.
- **Nearest neighbor + 1** connects a cells or cortical column with its 4 nearest adjacent cells or cortical columns. This is used to connect BCs to ACs and reciprocally.
- **Gaussian** connectivity connects the pre- and post-synaptic cell (column) with a weight proportional to a Gaussian function of the distance $d_{[pre,post]}$ between the pre- and post-synaptic cell (eq. (11) in the text).

References

Angelucci, A., Levitt, J., Walton, E., Hupe, J., Bullier, J., and Lund, J. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, 22(19):8633–8646.

Aquilué-Llorens, D., Goldman, J. S., and Destexhe, A. (2024). High-Density Exploration of Activity States in a Multi-Area Brain Model. *Neuroinformatics*, 22(1):75–87.

Benvenuti, G., Chemla, S., Boonman, A., Masson, G., and Chavane, F. (2015). Anticipation of an approaching bar by neuronal populations in awake monkey v1. *Journal of Vision*.

Benvenuti, G., Chemla, S., Boonman, A., Perrinet, L., Masson, G. S., and Chavane, F. (2020). Anticipatory responses along motion trajectories in awake monkey area v1. *bioRxiv*.

Berry, M., Brivanlou, I., Jordan, T., and Meister, M. (1999). Anticipation of moving stimuli by the retina. *Nature*, 398(6725):334—338.

Carlu, M., Chehab, O., Dalla Porta, L., Depannemaecker, D., Héricé, C., Jedynak, M., Köksal Ersöz, E., Muratore, P., Souihel, S., Capone, C., Zerlaut, Y., Destexhe, A., and Di Volo, M. (2020). A mean-field approach to the dynamics of networks of complex neurons, from nonlinear Integrate-and-Fire to Hodgkin-Huxley models. *Journal of Neurophysiology*, 123(3):1042–1051.

Cessac, B. (2022). Retinal processing: Insights from mathematical modelling. *Journal of Imaging*, 8(1).

Chemla, S. (2010). *A biophysical cortical column model for optical signal analysis*. PhD thesis, School of Information and Communication Sciences.

Chemla, S., Chavane, F., Vieville, T., and Kornprobst, P. (2007). Biophysical cortical column model for optical signal analysis. In Holmes, W. R., Jung, R., and Skinner, F., editors, *Sixteenth Annual Computational Neuroscience Meeting (CNS)*, volume 8, Suppl 2 of *BMC Neuroscience*.

Chemla, S., Reynaud, A., di Volo, M., Zerlaut, Y., Perrinet, L., Destexhe, A., and Chavane, F. (2019). Suppressive traveling waves shape representations of illusory motion in primary visual cortex of awake primate. *Journal of Neuroscience*, 39(22):4282–4298.

Chen, E. Y., Marre, O., Fisher, C., Schwartz, G., Levy, J., da Silviera, R. A., and Berry, M. (2013). Alert response to motion onset in the retina. *Journal of Neuroscience*, 33(1):120–132.

Despotović, D., Joffrois, C., Marre, O., and M.Chalk (2024). Encoding surprise by retinal ganglion cells. *PLOS Computational Biology*, 20(4):1–20.

Destexhe, A. (2009). Self-sustained asynchronous irregular states and up-down states in thalamic, cortical and thalamocortical networks of nonlinear integrate-and-fire neurons. *J Comput Neurosci.*, 27(3):493–506.

Di Volo, M., Romagnoni, A., Capone, C., and Destexhe, A. (2019). Biologically realistic mean-field models of conductance-based networks of spiking neurons with adaptation. *Neural Computation*, 31(4):653–680.

DiVolo, M. and Férézou, I. M. (2021). Nonlinear collision between propagating waves in mouse somatosensory cortex. *Scientific Reports*, 11:19630.

DiVolo, M., Romagnoni, A., Capone, C., and Destexhe, A. (2019). Biologically Realistic Mean-Field Models of Conductance-Based Networks of Spiking Neurons with Adaptation. *Neural Computation*, 31(4):653–680.

Ebert, S. (2023). *Dynamical synapses in the retinal network*. Phd thesis, Université Côte d'Azur.

Ebert, S., Buffet, T., Sermet, B., Marre, O., and Cessac, B. (2024). Temporal pattern recognition in retinal ganglion cells is mediated by dynamical inhibitory synapses. *Nature Communication*, 15:6118.

ElBoustani, S. and Destexhe, A. (2009). A master equation formalism for macroscopic modeling of asynchronous irregular activity states. *Neural computation*, 21(1):46–100.

Emonet, J. (2024). *A retino-cortical model to study movement-generated waves in the visual system*. Computer science, Nice Côte d'Azur, Valbonne.

Guo, K., Robertson, R., Pulgarin, M., Nevado, A., Panzeri, S., Thiele, A., and Young, P. (2007). Spatio-temporal prediction and inference by v1 neurons. *European Journal of Neuroscience*, 26:1045–1054.

Jancke, D., Erlaghen, W., Schöner, G., and Dinse, H. (2004). Shorter latencies for motion trajectories than for flashes in population responses of primary visual cortex. *Journal of Physiology*, 556:971–982.

Johnston, J. and Lagnado, L. (2015). General features of the retinal connectome determine the computation of motion anticipation. *Elife*.

Kartsaki, E., Hilgen, G., Sernagor, E., and Cessac, B. (2024). How does the inner retinal network shape the ganglion cells receptive field : a computational study. *Neural Computation*, 36(6):1041–1083.

Menz, M. D., Lee, D., and Baccus, S. A. (2020). Representations of the amacrine cell population underlying retinal motion anticipation. *bioRxiv*.

Overwiening, J., Tesler, F., Guarino, D., and Destexhe, A. (2024). A multi-scale study of thalamic state-dependent responsiveness. *bioRxiv*.

Papoulis, A. (1965). *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York.

Schwartz, G., Harris, R., Shrom, D., and Berry, M. J. (2007). Detection and prediction of periodic patterns by the retina. *Nature Neuroscience*, 10(5):552–554.

Sherman, S. and Guillery, R. (2011). Distinct functions for direct and transthalamic corticocortical connections. *J Neurophysiol.*, 106(3):1068–77.

Souihel, S. (2019). *Generic and specific computational principles for visual anticipation of motion trajectories*. Phd thesis, Université Nice Côte d'Azur ; EDSTIC.

Souihel, S. and Cessac, B. (2019). Anticipation in the retina and the primary visual cortex : towards an integrated retino-cortical model for motion processing. In *ICMNS 2019 - The 5th International Conference on Mathematical NeuroScience*, Copenhague, Denmark.

Souihel, S. and Cessac, B. (2021). On the potential role of lateral connectivity in retinal anticipation. *J. Math. Neurosc.*, 11(3).

Souihel, S., Cessac, B., Volo, M. D., Destexhe, A., Chavane, F., Chemla, S., and Marre, O. (2019). Anticipation in the retina and the primary visual cortex : towards an integrated retino-cortical model for motion processing. In *NeuroMod 2019 - First meeting of the NeuroMod Institute*, Fréjus, France.

Subramaniyan, M., Ecker, A. S., Patel, S. S., Cotton, R. J., Bethge, M., Pitkow, X., Berens, P., and Tolias, A. S. (2018). Faster processing of moving compared with flashed bars in awake macaque v1 provides a neural correlate of the flash lag illusion. *Journal of Neurophysiology*.

Veltz, R. (2011). An analytical method for computing Hopf bifurcation curves in neural field networks with space-dependent delays. *Comptes Rendus Mathematique*, 349:749–752.

Zerlaut, Y., Chemla, S., Chavane, F., and Destexhe, A. (2018). Modeling mesoscopic cortical dynamics using a mean-field model of conductance-based networks of adaptive exponential integrate-and-fire neurons. *Journal of Computational Neuroscience*.

6 Supplementary Material

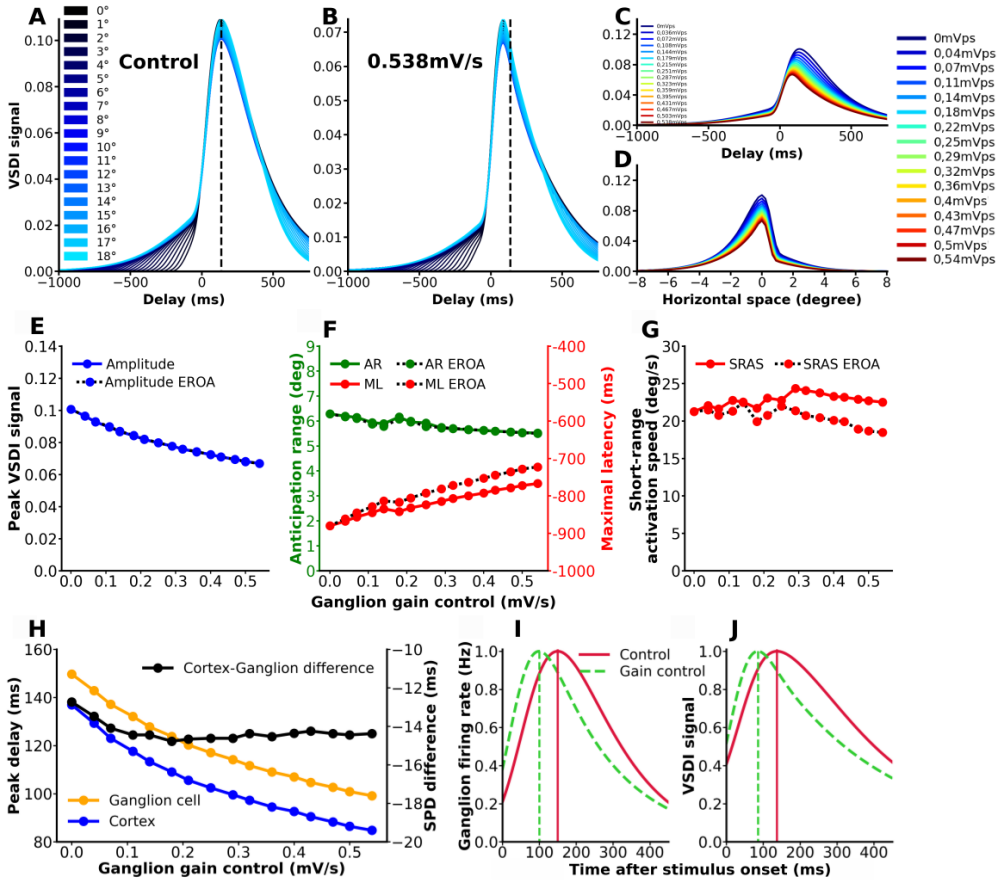

Figure S1: **The effect of RGCs gain control strength, h_G, on the cortical response. VSDI signal response A) without ganglion gain control (control) or at B)** 0.54**mV/s.** The color bar on the left of Fig. A represent the color gradient introduced in section 2.4. **C) Temporal VSDI signal** in response to increasing ganglion gain control for the cortical column located at the center of the layer ($x = 9°, y = 1.35°$). **D) Spatial VSDI signal** in response to increasing ganglion gain control for the time where the central cortical column reaches its maximum. The x coordinates has been shifted so that the central cortical column is actually located at $x = 0$. **E) VSDI signal amplitude** of the central cortical column versus h_G. **F) Temporal and spatial observables:** anticipation range (green) and maximal latency (red) versus h_G. **G) Speed observable :** short-range activation speed (red) in function of ganglion gain control weight. In A, B, C, we also drawn the Equivalent Retinal Output Amplitude (EROA) curve. This is the dotted black curve with the same coloured symbols. **H) Stationary peak delay (SPD)** for RGCs (orange), VSDI signal (blue) for the central cell (scales on the left) and difference between RGC SPD and VSDI signal SPD (black, scales on the right). **I) Shape of the central RGC response profile to the moving bar**, without gain control (red) and with RGCs gain control 0.54 mV/s (dashed green). Note that the two traces have been rescaled to have the same maximum. This is to emphasize the change in the shape of the response induced by RGCs gain control. **J) VSDI signal**, same conditions. In E,F, the dotted lines correspond to the peaks in the RGC firing rate or VSDI signal without RGCs gain control (red) and with it (green).

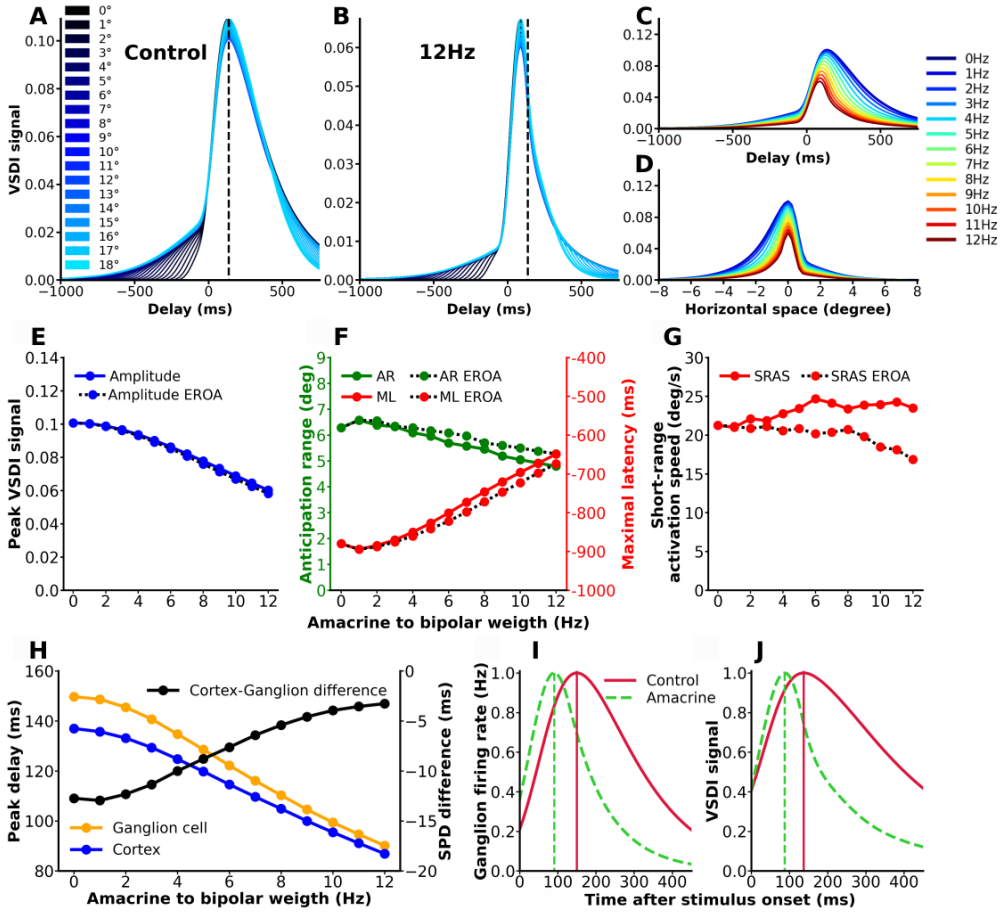

Figure S2: **The effect of ACs-BCs feedback loop on the cortical response. VSDI signal response A) without amacrine feedback inhibition (control) or at B)** 12**Hz.** The color bar on the left of Fig. A represent the color gradient introduced in section 2.4. **C) Temporal VSDI signal** in response to increasing macrine feedback inhibition for the cortical column located at the center of the layer ($x = 9°, y = 1.35°$). **D) Spatial VSDI signal** in response to increasing macrine feedback inhibition for the time where the central cortical column reaches its maximum. The x coordinates has been shifted so that the central cortical column is actually located at $x = 0$. **E) Amplitude quantity:** VSDI signal amplitude of the central cortical column versus the feedback loop weight. **F) Temporal and spatial observables:** anticipation range (green) and maximal latency (red). **G) Speed observable :** short-range activation speed (red) in function of amacrine to bipolar weight. **H) Stationary peak delay (SPD)** for RGCs (orange), cortical columns (blue) and difference between the cortical and RGC SPD (black). **I) Shape of the central RGC response profile to the moving bar**, without feedback AC connection ($w_B^A = -w_A^B = 0$) (red) and when these parameters take the maximum value ($\left| w_B^A \right| = w_A^B = 12$ Hz, dashed green). Note that the two traces have been rescaled to have the same maximum. **J) VSDI signal**, same conditions. The dotted lines in E,F correspond to the peaks in the RGC firing rate or VSDI signal without feedforward AC connection (red) and with it (green).